TECHNISCHE UNIVERSITÄT DARMSTADT

A Preliminary Version of a

Classical Unified Growth Theory

Simulation and Evaluation of an Economic Theory of Production per Capita

based on the Malthusian Theory of Population and the Smithian Theory of Production

Dissertation

Submitted in fulfillment of the requirements for the degree of

Doctor rerum politicarum (Dr. rer. pol.)

to the Department of Law and Economics TU Darmstadt

by

Diplom-Volkswirt Tim Lueger, MSQE

(Place of birth: Hann Muenden)

First supervisor: **Prof. Dr. Volker Caspari**
Second supervisor: **Prof. Dr. Jens Krueger**

Darmstadt 2019

Lueger, Tim: Classical Unified Growth Theory

Darmstadt, Technische Universität Darmstadt,

Jahr der Veröffentlichung der Dissertation auf TUprints: 2019

Tag der mündlichen Prüfung: 21.11.2019

Abstract

Throughout the history of economic thought, there have been numerous attempts to model an early era of "Malthusian" economic stagnation as well as the transition to an era of economic development in one coherent framework, or, in other words, a unified growth theory. In recent years, unified growth models have attracted a large readership among economists, challenging the conventional exogenous neoclassical growth theory. However, in most of these models, an important effect suggested by Malthus has been frequently omitted. By including what he had called "the great preventive check" in the conventional Malthusian trap model, which is based on the principle of population, the principle of diminishing returns and the principle of labor division, the transition can be modeled in a very simple dynamic macroeconomic framework. The correspondingly advanced theory suggests that increasing life expectancy tends to create a demographic structure that is much less prone to overpopulation. This new interpretation of the classical growth model is suggested to be capable of integrating the mechanisms of economic stagnation and economic development. Although the "vaguer intuitions" of the classical economists provided deeper and more profound insights than those of most modern unified growth theorists, the verbal form of their arguments has at the same time tended to be more favorable to misinterpretations. It is the intention of this work to identify these misinterpretations and to restore the main ideas of classical economics by building a basic classical unified growth model.

Zusammenfassung

In der Geschichte des ökonomischen Denkens gab es zahlreiche Versuche, eine Ära "Malthusianischer" ökonomischer Stagnation sowie ihren Übergang zu einer Ära ökonomischer Entwicklung in ein und derselben kohärenten Darstellung zu modellieren, oder in anderen Worten, in einer ganzheitlichen Wachstumstheorie (Unified Growth Theory). In den vergangenen Jahren haben Unified Growth Modelle eine zunehmende Leserschaft auf sich gezogen. In den meisten dieser Modelle wurde allerdings ein wichtiger klassischer Argumentationspunkt ausgelassen. Bezieht man die Auswirkungen dieses – von Malthus als "großen präventiven Check" (great preventive check) bezeichneten – Effektes in das konventionelle Malthusianische Modell, basierend auf dem Bevölkerungsprinzip, dem Prinzip der abnehmenden Grenzerträge und dem Prinzip der Arbeitsteilung, mit ein, so lässt sich der Übergang in einer sehr einfachen dynamischen, makroökonomischen Darstellung illustrieren. Die zu Grunde liegende Theorie besagt, dass ein Anstieg der durchschnittlichen Lebenserwartung eine demographische Struktur hervorruft, welche die Gefahr der Überbevölkerung zu großen Teilen eindämmt. Obwohl die "vagen Intuitionen" der klassischen Ökonomen tiefere und profundere Einsichten als die der meisten modernen Wachstumstheoretiker lieferten, war die verbale Form ihrer Argumente häufiger anfällig für Fehlinterpretationen. Die Intention dieser Arbeit liegt darin, diese Fehlinterpretationen mit Hilfe einer neuen Darstellung eines klassischen, ganzheitlichen Wachstumsmodells zu identifizieren und den wichtigsten klassischen Ideen wieder ihren rechtmäßigen Platz in der ökonomischen Wissenschaft zu sichern.

Preface

During my studies, I had numerous encounters with Unified Growth Theory in a wider sense. I firstly started to explore endogenous growth theory after realizing in a course on growth theory that neoclassical exogenous macroeconomic growth theory still appeared limited in modeling long–run growth. While I was always interested in a historical perspective, I discovered that most exceptional historical phenomena were rooted in economic causes when studying international economic history at Cardiff University. At Constance University, I was introduced to demographics in a seminar on population economics. As part of my focus on econometrics at the above universities, I also chose to analyze an econometric paper on economic stagnation employing pre–industrial demographic and economic time series. However, it was Clark's (2007) book "A Farewell to Alms" that made me rethink the entire process of human economic development, and having discussed Malthusian dynamics in a course on economic development, I was impressed by the idea of a population trap. Building on Galor's (2011) idea of a unified growth theory, my diploma thesis dealt with demography and wealth over the long run, trying to sustain the theoretical ideas of stagnation and development with empirical data. During the past ten years I was constantly seeking to maintain an inquisitive attitude, which led me to study numerous classical economic authors, who stressed the role of demographic changes. After I had reviewed the classical Malthusian dynamics, I realized that a consistent framework of the classical theory of population on which ecnomists can universally agree has not been established until today. To me this meant that either the theory lacks consistency or it has been misunderstood in important ways. With this dissertation I will argue that the latter is the case. Moreover, as I see the interplay between demographic change and population growth in developed and developing countries as the most urgent economic problem mankind is currently facing, I intend to design a classical mathematical model to enable projections of economic growth under different demographic circumstances to equip the science of economics with a more consistent understanding of the potential future economic development.

The objection might be raised that research of population development has traditionally been restricted to demographic science. However, an analysis of "overpopulation" needs to be based on economic observations, as it is concerned with the ratio between production and population. Historically, the interactions between population and economic production have attracted scientific interest for many centuries. Political considerations certainly contributed to suppress "the population question" from fully entering economic theory during the twentieth century. However, political correctness should not prevent the scientist from engaging in the broader taks of individual or collective "progress":

> Employed as he is upon a science, in which error, or even ignorance, may be productive of such intense and extensive mischief, he is bound, like a juryman, to give deliverance true according to the evidence, and to allow neither sympathy with indigence, nor disgust at profusion or at avarice; neither reverence for existing institutions, nor disgust at existing abuses; neither love of popularity, nor of paradox, nor of system, to deter him from stating what he believes to be the facts, or from drawing from those facts what appear to be the legitimate conclusions.[1]

Unfortunately, there exists no automatism that guarantees continuous scientific progress in the fields of economics and demographics as time advances, for every generation has to acquire anew the knowledge the former had accumulated by its own experience while facing very different environments. If we are lucky enough, we might encounter a historical figure in our studies who had the same deep thoughts many years before. The best way by which the scientist can try to improve the transmission of ideas between generations is by most transparently depicting his theory after having convincingly synthesized the body of literature of all former theories. Although this idealistic view might be wishful thinking, there should at least be a realistic effort to progress toward generalization and not to advance specialization and therefore fragmentation of the science, since knowledge is ultimately useless if it cannot be communicated between diferent scientific fields. In the case of fragmentation, ingenious methods developed from thousands of different viewpoints would be forgotten as soon as they cease to be useful and millions of unread articles will pile up without having ever been read. In the case of generalization, we may be able to unite and structure our collected knowledge in the most transparent and most accessible way.

In writing this dissertation, I received inspiration and assistance throughout the entire economic profession. I experienced great support from my teachers, colleagues und fellow students at the universities of Constance, Cardiff, Frankfurt and Darmstadt as well as at the IfW Kiel. First and foremost, I am deeply grateful to my supervisor and doctoral advisor Volker Caspari for his overall support, without which this project certainly could not have been completed within the past three years. I also thank Jens Krueger for his support as my second supervisor. Moreover, I am greatly indebted to Heinrich Ursprung, Ralf Brueggemann and Derek Matthews for their decisively inspiring suggestions. This work also benefited from extensive discussions with Christian Berker and Guenther Rehme. Besides, I would like to express my thanks to Uwe Hasser, Balint Tatar, Sabine Eschenhof-Kammer as well as an anonymous econometric reviewer for all the valuable support in methodological issues. Futhermore, I am indebted to Uwe Cantner, Oded Galor and an anonymous economic historian for mindfully assessing the relevant discussion papers, Bernard Beaudreau, John Bernell and Michael Neugart for important feedback on my presentations as well as Benjamin Friedman, Olivier de la Grandeville, Bertram Schefold and David Weil for very helpful theoretical comments. The completion of this dissertation was strongly accelerated by proofreading and administrative support from Casian Bardeanu, Ines Balta and Philipp Savage. I want to thank my parents for the unconditional trust and support during all the years of education. Finally, an especially heartfelt expression of gratitude is owed to Noélle-Christin for her loving, trust and patience.

[1] Senior (1836), p. 130.

Contents

List of Tables

List of Figures

List of Abbreviations

AD	Anno Domini
BC	Before Christ
GDP	real Gross Domestic Product
OLS	Ordinary Least Squares Method
PoDR	Principle of Diminishing Returns
PoG	Principle of Generation
PoLD	Principle of Labor Division
PoP	Principle of Population
USD	US-Dollars
VAR	Vectorautoregression

Chapter 1

Introduction: Unified Growth Theory

1.1 Overview

The aim of this dissertation is to construct and validate a unified model of economic growth that can explain the stylized (crude) evolution of GDP per capita over the very long run, i.e. over the complete history of mankind, in a single coherent model. It proceeds as follows.

In the introductory chapter, the evolution of GDP per capita will be illustrated and the two most prominent unified growth models will be presented.

The first part of the work (chapter two and three) deals with early human economic development. In chapter two, we will empirically evaluate the two models and conclude that there existed a regime of economic stagnation in Britain until around 1800 AD. Since the model of stagnation is chiefly based on a mechanism that had already been described by the classical economist T. R. Malthus in the year 1798, it is called "the Malthusian trap." In chapter three, this mechanism will be traced back to Malthus' and other classical economists' original understanding, which made use of three universal principles. On this basis, the plausibility of the "classical Malthusian trap" will be illustrated by the Darwinian process of evolution. In order to arrive at an empirically testable macroeconomic growth model, the principles will first be translated qualitatively into causal relationships. Then, these relationships will be quantitatively defined in a system of linear equations, exemplarily calibrated and simulated to show that the classical model can indeed account for the regime of stagnation.

Part II of the work (chapter four and five) proceeds with the more recent human economic development. In chapter four, the stylized facts of the second regime of economic development will be analyzed and the classical model of chapter three will be used to explain this process of development. Here, it will also be shown that the most popular contemporary approaches cannot explain the "escape from the Malthusian trap." As a promising alternative, this work will advance a fourth classical principle in chapter five, by which the switch from a regime of stagnation to a regime of development can be explained.

Part III (chapter six, seven and eight) unites the findings of the former two parts into one unified growth theory. In chapter six, the mechanism of the Malthusian trap as well as the escape from the Malthusian trap will be simulated in one and the same "unified growth model." Having checked the validity of the classical model with regard to the stylized facts, chapters seven and eight will deal with the separate empirical identification of each principle on which the classical unified growth theory rests. To this end, we will employ conventional OLS–estimations and a VAR–framework using country–specific annual historical data on crude birth rate, crude death rate and GDP per capita growth rate. The employed time series are mainly based on Mitchells (2013) "International Historical Statistics" and World Bank (2018) data.

From the results of Part III it will be concluded in Part IV that a reduction in mortality is generally capable of mitigating the rate of population growth responsible for stagnation and is a necessary condition for economic development. Consequently, it is argued that emerging economies follow a universal macroeconomic pattern of development. A decreasing death rate is succeeded by a decreasing birth rate, which at the same time induces GDP per capita to rise sustainably. Based on this finding, we will speculate about the future evolution of mortality and its consequence for economic development. The work concludes with the insight that the economic principles classical growth theory was built upon are found to prevail universally and that the explanatory power of classical growth theory remains, even after two hundred years, superior to most recent unified growth theories.

1.2 Evolution of GDP per Capita

It will certainly not be rated as an audacious statement to announce that the average person on Earth enjoys, in the year 2019, the highest material living conditions that have ever been recorded. Over the past two hundred years the world has experienced, albeit very unevenly distributed across countries and regions, unprecedented growth rates in terms of real gross domestic product (GDP) per capita. Maddison (2006) estimated that in the year 1830 the value of all goods produced amounted to about USD 700 billion, measured in 1990 USD. In the same year, the world population amounted to approximately 1.1 billion inhabitants. By the year 2003, population had grown to around 6.4 billion people, while GDP per capita had risen tenfold from approximately USD 640 to USD 6,400. Nonetheless, there is growing evidence that this most spectacular success story in the history of mankind is currently fading out, with the world economy gradually moving toward a stationary state, which some authors have already denoted as the return to "secular stagnation".[1] It may, therefore, be worthwhile to pause for a moment, to take stock of what has happened and to review and reassess this unique period of growth with the benefit of hindsight.

It is the general task of the growth economist to analyze and to explain "economic development over the very long run", i.e. to reveal causes for changes in GDP per capita from the dawn of history until today. Although the reconstruction of historical data determining past living conditions has always been and will always be subject to some debate, it is obvious that GDP per capita cannot have increased over the past few thousand years at the same speed as it did over the past two hundred years. A re–projection of those growth rates would result in "absurdly low"

[1] See, for example, Cervellati et al. (2017a).

living conditions during medieval times.[2] Consequently, it is plausible to presume a pre–modern era of stagnation or at least very slow growth that J. M. Keynes[3] (1930) had described as follows:

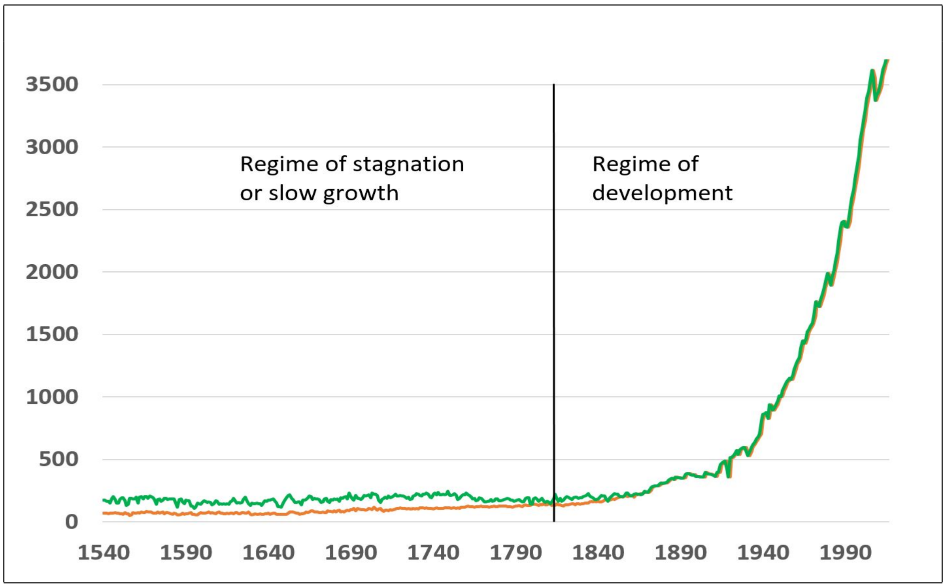
> From the earliest times of which we have record–back, say, to two thousand years before Christ — down to the beginning of the eighteenth century, there was no very great change in the standard of life of the average man living in the civilised centres of the earth.[4]

By that time, he was well aware of the fact that – roughly since the year 1800 – his countrymen had experienced a transitional phase from economic stagnation to growth, optimistically concluding that

> assuming no important wars and no important increase in population, the economic problem may be solved, or be at least within sight of solution, within a hundred years.[5]

Among others, Clark (2009) and Broadberry et al. (2015) seized on the "economic problem" and collected British historical data of GDP per capita to provide evidence of the transition from a historical regime of stagnation or very slow growth to a regime of development in the form of the popular "hockey stick", as is depicted in Figure 1.1.

Figure 1.1: Stylized facts of stagnation and development.



Sources: Broadberry et al. (2015), Clark (2009) and Mitchell (2013).

These data are often assumed to not simply reflect British economic history, but to represent a global development pattern in so far as every economy once found itself or is currently located in

[2] See Mokyr and Voth (2010). A re–projection of the initial example would thus suggest a GDP per capita value of approximately USD 60 in the year 1650.

[3] Sir John Maynard Keynes (1883–1946), British economist, member of the Royal Commission in 1913, financial representative for the Treasury to the 1919 Versailles peace conference, founder of modern macroeconomics.

[4] Keynes (1930), p. 1.

[5] ibid., p. 4.

a regime of stagnation or slow development. Moreover, the data suggest that something changed around the year 1800, when Britain apparently became the first economy to "mircaculously" generate sustained economic development.[6] As Broadberry and O'Rourke (2010) put it,

> viewed in the grand sweep of history, this change was undoubtedly radical, and must be ranked alongside other epoch–making changes such as the change from hunting and gathering to settled agriculture.[7]

Reluctantly during the first hundred years, then progressively catching up, the major part of the world economy succeeded the British example.

When staring at these facts, almost automatically some questions come to mind. "Why was there economic stagnation for so long?" "Why have some economies experienced development while others are still caught in stagnation?" "Will this improvement in material well-being be a lasting one?" Or in summary: "Why and when does long–run economic development occur?" With Lucas (1988) arguing that "the consequences for human welfare involved in questions like these are simply staggering"[8] and North (2013) urging that the elucidation of the transition to economic development seems to be "the most important historical question that might conceivably be possible to answer", the primary object of this work is to analyze the effects that made for an era of stagnation or slow growth as well as those enabling the transition to growth and thus to build a "very long–run" economic growth model, or to use Keynes' wording, to solve the economic problem. If we are able to explain the past, we may also be better able to assess the potential future development.

1.3 Unified Growth Theory

1.3.1 The demographic view on unified growth theory

As a starting point for theoretical analyses of GDP per capita growth, economists usually refer to the conventional neoclassical Solow (1956) growth model, for which Solow was awarded the Nobel Memorial Prize in Economic Sciences in the year 1987 due to "his contributions to the theory of economic growth". However, as this model appeared to be merely intended to describe long–run growth during the twentieth century, numerous authors began to look beyond its limitations, attempting to build a framework that could encompass the complete history of mankind to account for the "very long run". Mokyr and Voth (2010) summarized the emergence of this theoretical literature by stating that

> From the 1990s onwards, scholars started to search for an overarching theory that could encompass both slow growth and the transition to rapidly increasing per capita incomes — a "unified growth model". The field has flourished since. A number of themes stand out — demography, the influence of institutions, human capital and culture, and the role of technology.[9]

[6] The date 1800 is often chosen as a rough estimate to mark the British "take–off".
[7] Broadberry & O'Rourke (2010), p. 1.
[8] "Once one starts to think about them, it is hard to think of anything else." Lucas (1988), p. 5.
[9] Mokyr and Voth (2010), p. 8.

Firstly, as is noted in the above quote, a unified growth theory ought to be able to connect an early historical regime of stagnation or slow growth with a more recent historical regime of growth in productivity.[10] Secondly, a unified growth theory is expected to explain economic development endogenously, i.e. from within the model.

1.3.2 Endogenous unified growth theory

Over the past 25 years, countless theories in the spirit of Romer's (1986) growth theory of endogenous "technological change" have been developed by growth economists with the intention to explain the observed economic development. Out of Mokyr and Voth's above list, this work will support the first, the demographic view on unified growth theory and can indeed be declared as one long argument in favor of the deep virtues of population dynamics as a decisive component of very–long run economic growth.[11] Given the – in the following briefly reviewed – recent literature on endogenous growth theory, the focus on demographic variables is certainly not surprising. Notable merits of this view are that records of demographic data are among the most reliable, objective and transparent historical sources, relatively easy to collect, and arguably possess the longest historical time series. Furthermore, demographics might be most appropriately considered as a "hard" social science, since their index numbers do not leave much room for interpretation.

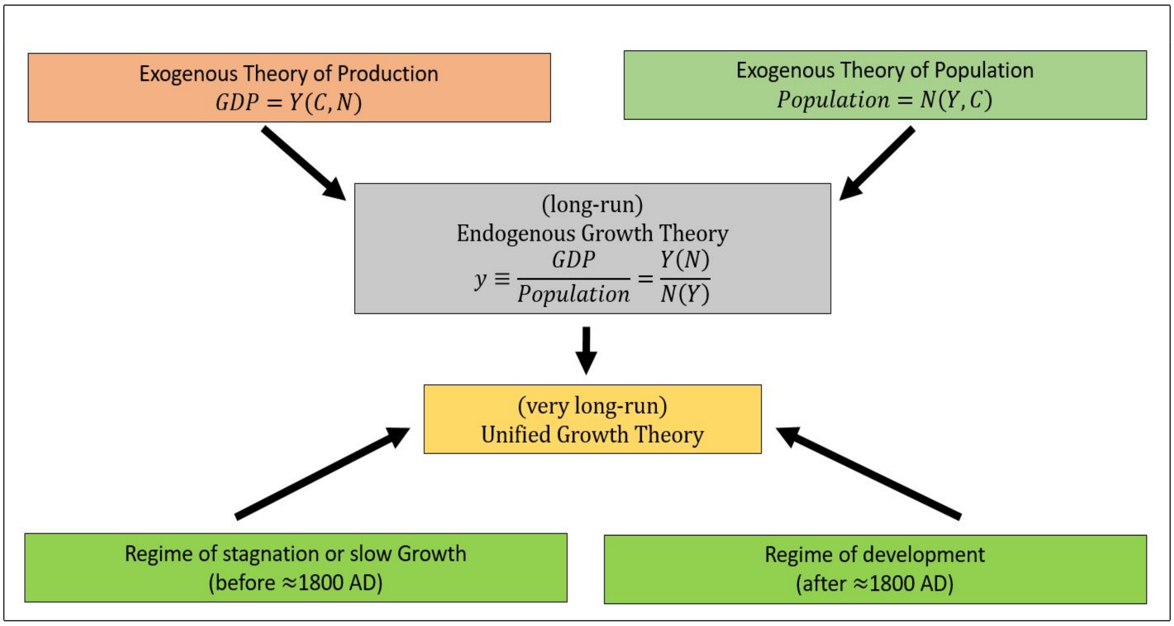
More importantly, as real GDP per capita (in the following denoted as "y") is defined as the ratio of real gross (total) production and population size (in the following simply denoted as "production" ("Y") and "population" ("N")), it merely reflects the productivity of the production factor population $\frac{Y}{N}$ and must be considered as a function of the two variables. On the one hand, the denominator simply reflects the degree of abundance or scarcity of population. On the other hand, the numerator production incorporates the value of all produced wealth within a given territory in a given period and is again a function of all available production factors ("C") in addition to population. It does not require much abstraction to envision that the existence of a population requires a certain amount of production and that the latter requires a population to be produced, i.e. the one cannot exist without the other. A synthesis of these two elements, where production may be determined by population and vice versa, allows us to call this mutual relationship $y(Y,N) = \frac{Y(N)}{N(Y)}$ an endogenous (demographic) growth theory. The first component of such an endogenous growth theory – a theory of production – seeks to model the effect of population on production. The second component is expected to explain – in contrast to the traditional exogenous Solow model – the causes for population growth by some other variable from within the model, ideally by production. Such an explanation will be referred to as a

[10] As a simplification, long run models often assume production per capita, income per capita and average living standards to be identical. Furthermore, production per capita will be abbreviated as *productivity* (of the production factor population). We will also, following Lucas (1988), refer to an increase in productivity (intensive growth) as *economic development*, whereas the expression *economic growth* will be used to characterize an increase in gross production (extensive growth). These terms have often been confused in the past, in particular in that literature which is concerned with "growth in economic development."

[11] "The great events of history are often due to secular changes in the growth of population and other fundamental economic causes, which, escaping by their gradual character the notice of contemporary observers, are attributed to the follies of statesmen or the fanaticism of atheists." Keynes (1919), chapter I.

theory of population. In summary, the demographic view on unified growth theory is illustrated in Figure 1.2.

Figure 1.2: The Demographic View on Economic Growth.



Consequently, most growth economists seem to accept some mechanism of a mutual relationship between population and production. However, they often disagree on the corresponding growth rates g_i that are related linearly in the following way:[12]

$$g_y = g_Y(g_N) - g_N(g_Y)$$

With regard to the demographic theory of production, two opposing lines of thought can be identified until this day, which might be marked as the "optimist" and the "pessimist" view on the "population question."[13] While the optimists claim that an increase in population growth will – chiefly owing to concomitant specialization and technological progress – raise average production per capita ($g_Y(g_N) > g_N(g_Y)$), the pessimists maintain that productivity would decline as a result of resources becoming relatively more scarce ($g_Y(g_N) < g_N(g_Y)$). To provide examples, we will briefly contrast the two most prominent unified growth theories.

---

[12] $g_y \equiv \frac{y_t - y_{t-1}}{y_{t-1}} \equiv \frac{\frac{Y_t}{N_t} - \frac{Y_{t-1}}{N_{t-1}}}{\frac{Y_{t-1}}{N_{t-1}}} = \frac{Y_t N_{t-1}}{N_t Y_{t-1}} - 1 = \frac{(1+g_Y)}{(1+g_N)} - 1 \Leftrightarrow ln(1+g_y) = ln(1+g_Y) - ln(1+g_N) \Leftrightarrow g_y \approx g_Y - g_N$.

The approximation holds for small growth rates (as rule of thumb, growth rates should not be larger than 10% such that the deviation from the true growth rate amounts to less than 5%).

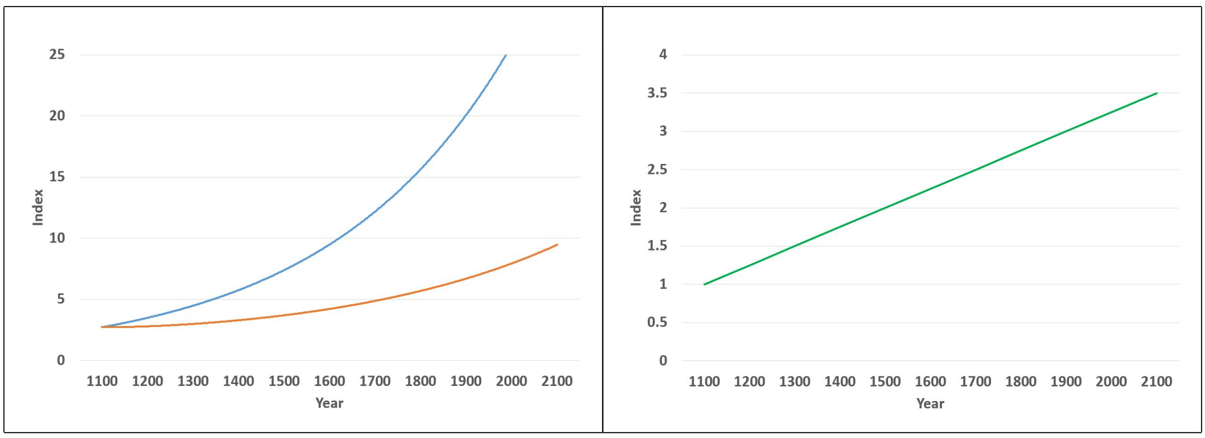
[13] See, for example, Bloom et al. (2003). For an optimist view, see, for example, Becker (1988) or Boserup (1965). Examples of a pessimist view are Hardin (1968) or Ehrlich (1968).

1.3.3 A unified growth theory consisting of one regime

Although theoretical considerations on the interaction between population and production are older than the science of economics itself, Kremer (1993) seems to have been one of the first authors to advance a mathematical framework for an endogenous unified growth theory. Following the demographic view, the two variables production and population in his model are crucial to roughly describing the historical evolution of GDP per capita. In a nutshell, his theory of production states that a larger population would generate a larger number of "technological" improvements, which would in turn increasingly upgrade to efficiency of some other production factor and thus raise production.[14] As for his theory of population, Kremer intuitively assumed that growth in production would in turn tend to raise population. The resulting parallel movement of population and production can be modeled by assuming an exponential increase in both variables as is depicted in the left graph of Figure 1.3. Kremer suggested a virtuous cycle between both variables, by which production would constantly outgrow population and therefore lead to a long gradual increase in GDP per capita since primitive times (see right graph of Figure 1.3).[15] The stylized facts of the model can be mathematically summarized as follows:

$$g_y = g_Y(g_N) - g_N(g_Y) > 0 \; with \; \frac{\partial g_Y}{\partial g_N} > 0, \frac{\partial g_N}{\partial g_Y} > 0.$$

Figure 1.3: Stylized evolution of production (left, blue), population (left, orange) and productivity (right, green) over the very long run according to Kremer (1993).



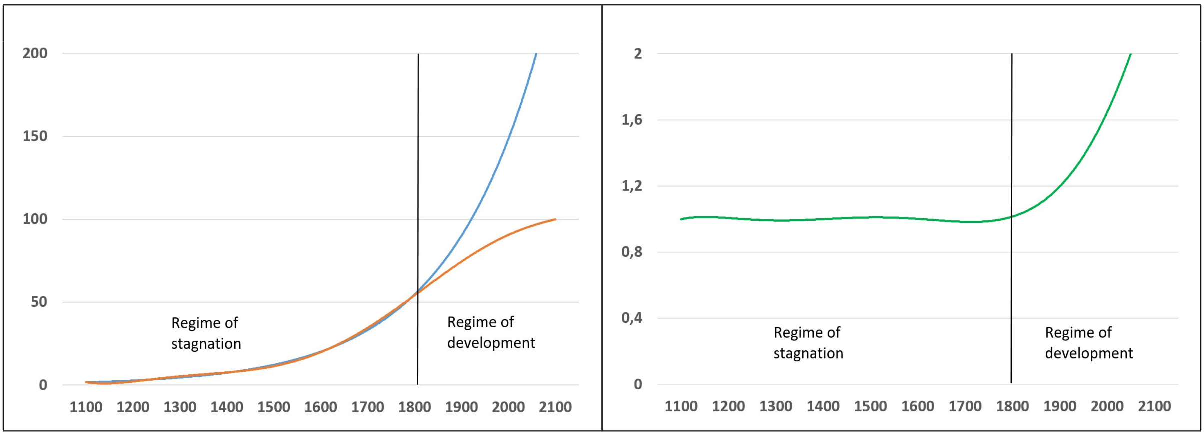
[14] To the economic historian, increasing GDP per capita is often depicted as the result of improvements in "technology". While economic theory still lacks a precise definition of technology, the usual broad definition includes productivity improvements due to factors like culture, institutions, enlightenment etc. Kremer (1993), for example, uses positive externalities from knowledge spillovers to model this effect, others emphasize increasing returns of population from a division of labor. We wil return to the problem of "technology" in the third and fourth chapters.

[15] "Early models in unified growth theory, such as Kremer's paper, modelled the transition from stagnation to growth as one long, gradual acceleration of [per capita] growth rates. [...] Kremer's model assumes that more people spell faster technological change, since the probability of a person having a bright idea is more or less constant. [...] Since 1,000,000 BC, growth rates of population can be predicted from the current size of the population." Mokyr and Voth (2010), p. 8.

1.3.4 A unified growth theory consisting of two regimes

In contrast to the view of a slow gradual evolution over the whole time span under observation, more recent unified growth models emphasize the more differentiated view of two regimes of development: One regime of stagnating GDP per capita until about 1800 AD and another regime of economic development thereafter.[16] Galor and Weil (1999), for example, challenged the conventional Kremerian view by adding that production growth did not succeed in outperforming population growth until about 1800 AD and that the latter would eventually fully consume any productivity gains from "technological" improvements. In addition, since economic development had obviously begun by the end of the nineteenth century, they extended their model of stagnation by suggesting that the positive effect of production on population broke down after 1800 AD and that a slowdown in population growth became a main determinant of the simultaneously observed growth in GDP per capita, whereas production continued to grow in the same manner as before.[17] The corresponding stylized facts are illustrated by Figure 1.4.

Figure 1.4: Stylized evolution of production (left, blue), population (left, orange) and productivity (right, green) over the very long run according to Galor and Weil (1999).



As a result, the regime of stagnation might be summarized by writing

$$g_y = g_Y(g_N) - g_N(g_Y) = 0, \ with \ \frac{\partial g_Y}{\partial g_N} > 0, \frac{\partial g_N}{\partial g_Y} > 0,$$

while the regime of development may correspond to the following formulation

$$g_y = g_Y(g_N) - g_N(g_Y) > 0 \ with \ \frac{\partial g_Y}{\partial g_N} > 0, \frac{\partial g_N}{\partial g_Y} = 0.$$

Although, therefore, population is in both frameworks regarded as a positive endogenous source of production and vice versa, Kremer's theory spurs a cycle of economic development over the very long run (in the following denoted as "the cycle of prosperity"), while Galor and Weil

[16] See, for example, Hansen and Prescott (2002) and Tournemaine and Luangaram (2012).
[17] The main findings of Galor's research are summarized in his book "Unified Growth Theory" (Galor 2011).

suggest a cycle of economic stagnation that weakens after 1800 AD ("the cycle of misery").[18] To investigate which of the two views appears to be confirmed by the real economic development, this work will proceed chronologically and first analyze the interactions between demographic and economic variables during the early era until about 1800 AD (Part I) and afterwards examine the corresponding interactions during the late era of development after approximately 1800 AD (Part II).

[18] See Livi–Bacci (2012) for a historical description of this effect and for an overview of population history.

Part I

The Era of Economic Stagnation

Chapter 2

The Regime of Economic Stagnation

2.1 Evolution of GDP per Capita

2.1.1 The data – why Britain?

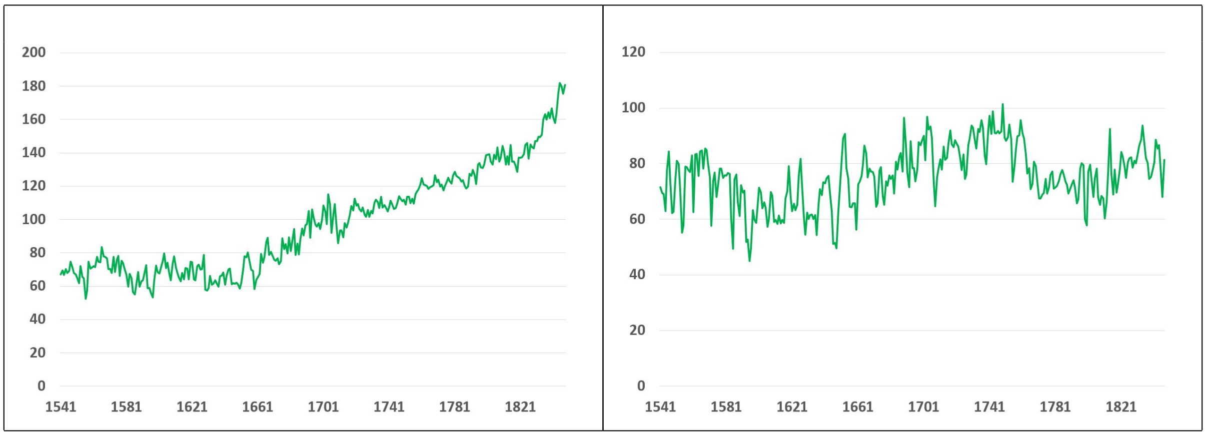
As in the introductory chapter, our investigation on the evolution of GDP per capita during the early regime will heavily focus on the British data.[1] This happends for the following reasons. Firstly, due to her early take–off and British global economic leadership during the nineteenth century, the country is often viewed as a globally representative case in economic development, attracting more interest among economists than any other. Secondly, the resulting discussions among economic historians have favoured the collection/reconstruction of a large number of British high–quality historical long–run data series related to economic development, more than for any other country. Thirdly, the extraordinarily long period of relative peace on the British Islands without much foreign intervention since the eleventh century allowed for a comparatively steady development without large breaks in the time series, i.e. for a "natural" development.

In spite of these merits, the pattern of the British historical GDP per capita evolution remains debated. To examine its evolution until the economic take–off during the nineteenth century, the currently most common collections of British data from Figure 1.1 are displayed on a shorter time scale in Figure 2.1, starting in 1541 and ending in the mid-nineteenth century. The left graph of the Figure illustrates Broadberry et al.'s (2015) GDP per capita data, the right graph shows Clark's (2009) data. In the latter case, although there exist pronounced fluctuations in living standards, a linear regression of the series does not display an obvious positive trend. Broadberry et al.'s data, on the other hand, exihibit, at least after the year 1650, an upward trend. While Broadberry et al.'s empirical calculations yield some support for Kremer's theoretical view of a slow gradual increase in GDP per capita, Clark's historical estimates suggest a period of economic stagnation, sustaining Galor's ideas with regard to a unified growth theory.[2] We will therefore now roughly evaluate the reliability of both data series and conclude that the Clark time series appears to be the much more plausible one.

[1] We will in the following crudely refer to "Britain" as the entity for which representative data are provided. A more differentiated use of the term would require a tedious examination of regional data.

[2] Clark even speculates that the average English person was in the year 1800 not better off than their ancestors on the African plains millennia before.

Figure 2.1: Indexed British GDP per capita 1541–1848. Left graph Broadberry et al. data, right graph Clark data.



Sources: Broadberry et al. (2015) and Clark (2009).

2.1.2 Quantitative evidence of early stagnation

First of all, it has to be kept in mind that the different pattern of the above series might have merely been the result of personal rivalries. In his 2007 book "A Farewell to Alms", Clark elaborated (although not mathematically) on a unified growth theory incorporating the two regimes mentioned in the last chapter. Since his main theses seemed to be ad odds with the at that time prevailing main source of historical data series of GDP per capita by Maddison (2006), Clark (2008) blamed – arguably deservedly – the Maddison data series for being predominantly random extrapolations and constructed his own series of British GDP per capita.[3] As a response and probably in avoidance of leaving the monopoly of historical interpretation to Clark, Broadberry et al. (2015) constructed another historical series of British GDP with the apparent aim of updating the Maddison Project Database. Their main argument was that Clark's data did not take into account the increasing number of working days of laborers over the course of the centuries. Clark (2018) naturally rejected these allegations.[4]

Since this work is more concernced with the stylized facts of economic development, giving much less weight to single historical instances, we will defer the debate on the detailed reconstruction of the data series in question and focus on their empirical plausibility. Firstly, Clark's critique that the Broadberry et al. series would imply implausibly low GDP per capita values during preindustrial times appears to be the most convincing argument for rejecting the latter series. Secondly, if we hypothetically accepted the Broadberry et al. data series as being the true one, the question of stagnation would merely be deferred to an earlier period. As we can see from Figure 2.1, stagnation exists in the Broadberry et al. data as well, merely some 150–200 years before the take–off in the Clark data. If we were to agree on this finding, there would be no question about the *existence* of two regimes, but rather a question about the exact timing of

[3] See Clark (2008).
[4] See Clark (2018).

the transition between the regimes of stagnation and development. Thirdly, in contrast to the two above depicted data series on GDP per capita, Allen's (2001) data series on real wages of London laborers is less suspicious of opportunism, as it was constructed well before the above debate had begun. The real wage, or the "value of labor", is a useful proxy variable for GDP per capita as, in growth theory, it is supposed to move proportionally with GDP per capita and often serves as an alternative measure of living standards. As can be seen from Figure 2.2, the trend of the real wage series seems to fit the pattern of the Clark data much better than that of the Broadberry et al. data in the sense that a significant increase in real wages cannot be observed until the year 1800. Livi–Bacci generalizes this finding by even stating that

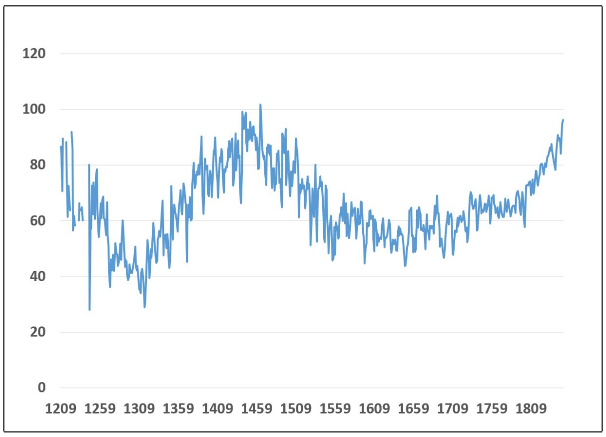
> real wages in general declined throughout Europe during the eighteenth century and into the first decades of the nineteenth.[5]

This view is hardly compatible with a slow permanent increase in GDP per capita as suggested by Broadberry et al. Fourthly, indicators regarding the pros and cons of stagnation or a gradual change include anthropometric and archaeological evidence. Even if Clark's estimates on GDP per capita and Allen's estimates on real wages were questioned, the generally observed stagnation of body stature due to insufficient nutrition until around 1800 AD would provide strong evidence of economic stagnation.[6] As Livi-Bacci continues,

> Another indication is variation in average height, which seems in this same period to have declined in England, in the Hapsburg Empire, and in Sweden.[7]

If we were to follow the Broadberry et al. dataset, we should instead suppose that increasing GDP per capita would be well–reflected by a parallel growth of body stature. Altogether, the evidence for a regime of economic stagnation seems to be stronger than the evidence for a regime of slow growth. We will continue this line of thought in the next section.

Figure 2.2: Indexed London real wages 1209-1848.



Source: Allen (2001).

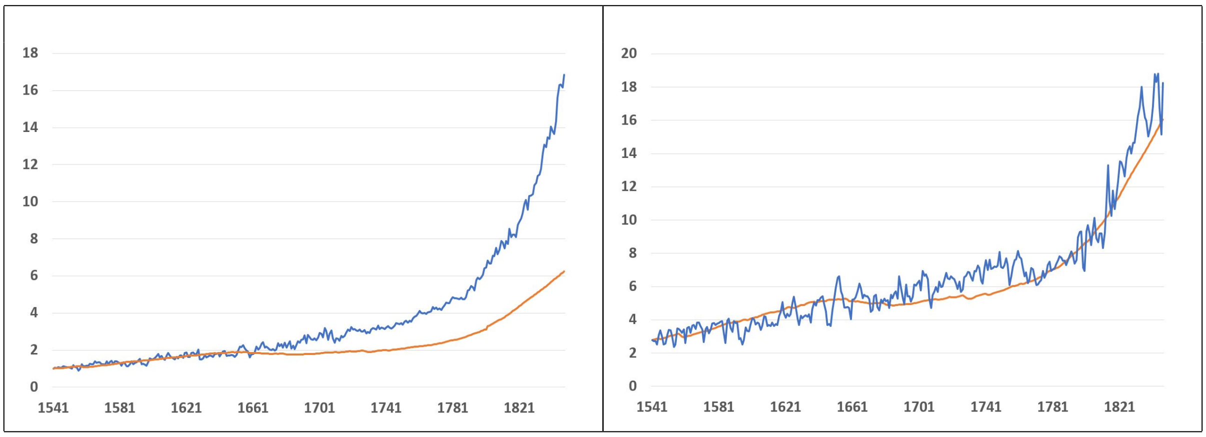
[5] Livi–Bacci (2012), p. 72.
[6] See Tanner (1994), or more recently in Hinde et al. (2018).
[7] Livi–Bacci (2012), p. 72.

2.2 Evolution of GDP and Population

2.2.1 Quantitative evidence of parallel growth in production and population

At first sight, the idea of premodern economic stagnation seems to contradict the intuition of the modern layman that “growth” followed a positive long–run trend. At second glance, however, we find that we have to distinguish between “economic growth” (in GDP) and “economic development” (in GDP per capita). Although the notion of stagnation might lead to the suspicion that production and population had stagnated as well until the mid–nineteenth century, Figure 2.3 refutes this view by illustrating the development of production and population separately, again based on the Broadberry et al. data (left) and the Clark data (right). Population estimates are in both cases based on the relatively well–established estimates of Wrigley and Schofield (1981) and accordingly not subject to major debate. Since both series uncover that the two variables not only increased over time, but also increased steadily and exponentially, the fact of continuous “growth” in production seems incontrovertible and with it the confirmation of the idea of continuous “technological progress”. However, the favored Clark data (right graph) again tend to support a “vicious cycle of misery” as advanced by Galor rather than a “virtuous cycle of prosperity” as suggested by Kremer. This implies that population growth used to “win the race” against growth in production, writing $g_N \geq g_Y$, leaving no room for economic development in terms of GDP per capita.

Figure 2.3: Indexed British rise in GDP and population size 1541–1848.



Sources: Broadberry et al. (2015), Clark (2009), Wrigley and Schofield (1981).

If we proceed to conjecture that the Clark data are a good description of reality and are representable for other economies, we ought to observe roughly proportional increases in GDP and population over the long run not only for Britain, but for every country. Consequently, we expect plenty of evidence from international comparisons. The observation of stagnation in some African states up to this day, in spite of strong simultaneous population growth therefore serves as a fifth argument in favour of Galor’s view. Moreover, the quite generally displayed historical constancy of productivity in economies with considerable total economic growth must be equally accredited to a rapidly growing population. Among the largest economies we find that the lowest levels of GDP per capita are regularly preceded by strong growth in both production and population, be it China (during the years 1959–1961), India (1943) or Russia (1932–1933), rejecting once again the possibility that production would steadily outgrow population. It seems on the contrary rather reasonable to categorize the current global situation of economies into two groups: Those economies still trapped in stagnation and those that escaped from a low level of stagnating productivity. Examples of both groups are depicted in Appendix 11.1.

Sixthly, past British economists were very well aware of their own economic situation in the nineteenth century and one of the best ways to find evidence of stagnation in early nineteenth century Britain is, of course, to adhere to the writings of the contemporaries who witnessed this stagnation. The time frame of Figure 2.1 has been chosen deliberately, since the ending year 1848 marks the latest possible date at which we may still talk about existing economic stagnation. Nonetheless, even as late as 1848, J. S. Mill[8] doubted the superior power of growth in production and "technological progress" as compared to growth in population:

> Hitherto it is questionable if all the mechanical inventions yet made have lightened the day's toil of any human being. They have enabled a greater population to live the same life of drudgery and imprisonment.[9]

Clearly, Broadberry et al.'s data series contradicts this assessment, insinuating that these authors had a better knowledge of the classical British economy than the classical economists themselves. This point will be explored more extensively when the precise mechanism of stagnation is presented in chapter three. Finally, as the presumption of population tending to outgrow production seems to be supported by the data, this view has found broad acceptance in the discipline of cliometrics.

> Cliometrics confirms that [the cycle of misery in] the preindustrial economy is a good description for much of demographic–economic history;[10]

In summary, since the (former) existence of a regime of stagnation in GDP per capita is sustained by sources with all kinds of evidence, we will eventually assume that production and population grew roughly proportionally until around 1800 AD.

[8] John Stuart Mill (1806–1873), British philosopher, Rector of the University of St Andrews, Member of Parliament for Westminster.
[9] Mill (1848), book IV, chapter VI.
[10] James Foreman-Peck (2019).

2.3 The Regime of Economic Stagnation: Stylized Facts

Nonetheless, we must remark that, when trying to generalize the British case on a global scale, it is reasonable to consider the term “stagnation” merely in its abstract form as a stylized fact and not to insist on the prevalence of a fixed productivity level in every single historical instance. There was certainly some sporadic economic development in the ancient Greek and Roman economies as well as in some regions of late medieval China, Europe or Japan. However, as a starting point for theoretical investigations on the occurrence of the transition to “modern exponential” development, the abstraction of stagnation will be extremely useful in contrasting the two regimes of a unified growth theory. Hence, the stylized facts of the regime of stagnation can be summarized as follows.

1. We observe “economic stagnation”: GDP per capita is constant over the long run:

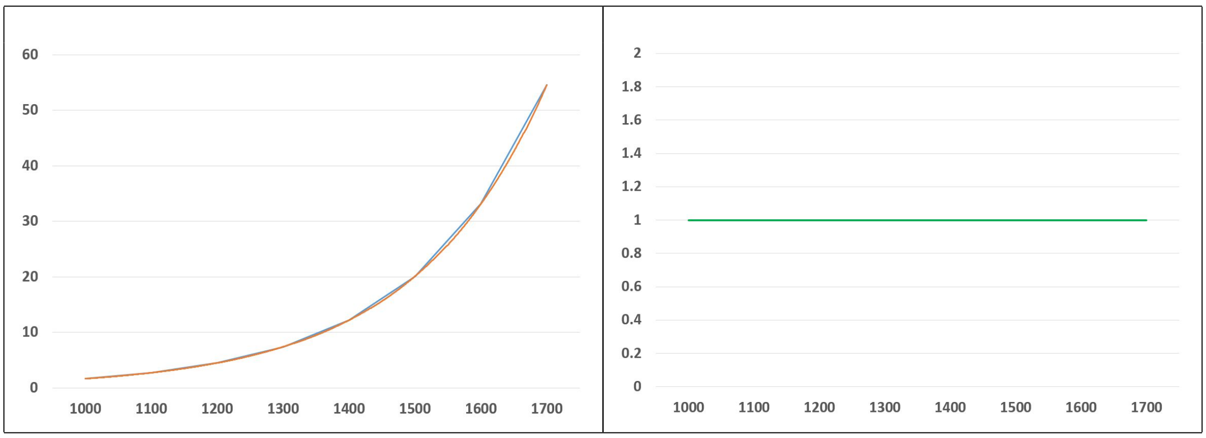
$$g_y = g_Y(g_N) - g_N(g_Y) = 0$$

2. We observe a “cycle of misery”: GDP and population grow roughly with the same positive rate over the long run:

$$\frac{\partial g_Y}{\partial g_N} = \frac{\partial g_N}{\partial g_Y} > 0$$

Since this stylized regime of stagnation has been found to have prevailed for at least four hundred years on a comparatively low level of production per capita, it may be properly called the “era of economic stagnation”. The stylized facts of the regime of stagnation are illustrated in Figure 2.4.

Figure 2.4: Left: The stylized fact of the “cycle of misery” of increasing population (left, orange) and production (left, orange) in Britain. Right: The stylized fact of stagnation in GDP per capita (green).



Chapter 3

A Classical Growth Theory of Stagnation

3.1 Advancing the Malthusian Interpretation of History

3.1.1 Galor's "Malthusian trap"

So far, we have merely employed descriptive statistics to follow the evolution of GDP, GDP per capita and population. In this chapter, we will investigate the theoretical mechanism yielding the "cycle of misery" between production and population more closely to account for an endogenous theory of growth. As we have seen, the historical evolution of the three variables seems to be roughly in line with the theory of stagnation advanced by Galor and Weil. As Galor and Weil borrowed their theory from the classical economist T. R. Malthus[1], they referred to the responsible mechanism as "Malthusian trap". Although the relatively neutral wording "population trap" is still widely employed as describing the theoretical mechanism of stagnation, a number of quite different labels have been circulating ranging from F. Lassalle's[2] (1863) "iron law of wages" to Keynes' (1930) "struggle for subsistence". Since all of these terms share the same theoretical fundament, namely the apparent inevitability of stagnating productivity under the prevalence of excessive population growth, they are in the following summarized under the now established expression among unified growth economists, the "Malthusian trap".

The "Malthusian trap" as a more detailed mechanism of the "cycle of misery" may be assumed to be owed to three effects that are described in equation 3.1. Firstly, Galor and Weil seem to have agreed with Kremer that population growth might trigger the invention of new technologies and therefore positively affect growth in production (1). Secondly, although it is not always explicitly stated, population growth is usually assumed to be somehow positively initiated by growth in production (2). Thirdly, Galor and Weil suggested that the higher growth rate in population seems to have steadily operated to neutralize its simultaneous advances in production in the

[1] Thomas Robert Malthus (1766–1834), British professor of history and political economy at the East India Company College in Haileybury, fellow of the Royal Society.

[2] Ferdinand Lassalle (1825–1864), writer, socialist politician and founding father of the German social democrat party.

British economy before 1800 AD, decreasing productivity and emphasizing diminishing returns to population (3). These three effects seem to have held productivity roughly on a constant "subsistence level" and are conjointly formulated by

$$g_y = \underbrace{\overbrace{g_Y(g_N)}^{(1)} - \overbrace{g_N(g_Y)}^{(2)}}_{(3)} = 0 \ with \ \frac{\partial g_Y}{\partial g_N} > 0, \frac{\partial g_N}{\partial g_Y} > 0. \tag{3.1}$$

Galor and Weil's research resuscitated – together with Clark's (2007) book – a debate that had lingered subliminally in economic thought for about two hundred years.[3] As a result, Malthusian ideas have attracted renewed interest among growth economists over the past two decades. As it appears to offer a simple framework in explaining economic long–run stagnation in productivity, some interpretation of a Malthusian trap is today commonly viewed as a cornerstone of a unified growth theory. However, notwithstanding their reference to Malthus, it appears that the fewest of those unified growth authors have actually read Malthus' original publications on the subject, as their theories often grossly contradict Malthus' original statements.[4] To make up for this defect, this work will advance the original view of classical economics, supplementing Malthus' theory with theories established by other classical and neoclassical economists. This retrospective view is considered to be necessary, since a thorough understanding of the mechanism of the Malthusian trap is both essential to perceive the social, macroeconomic scope of population growth and crucial in understanding the transition toward economic development.[5] Once we have established the classical view, we will find that understanding the answer to the question "why was there stagnation?" is even more illuminating to all social sciences than solving the puzzle of "why there was development." Once the principles responsible for stagnation are understood, only a small step remains to understanding the unified growth process.

3.1.2 Classical economics

Among others, North (2013) reminded us that the origins of the question of stagnation could be traced back to classical economic theory, which had already deeply influenced philosophy and natural sciences until the middle of the nineteenth century and whose agenda was not much different from that of current unified growth theory.

Over a full century, roughly ranging from 1770–1870, when economics was known as "political economy", demographics played a vital role in the theory of growth. The earlier mercantilist theory, facing regular devastating mortality crises, had viewed a large population as the fundament of (total) national economic prosperity in the international race for scarce resources (see, for example, Mun (1664)). Thereafter, the French economist and statesman Turgot (1767), witnessing the French population explosion, seems to have been one of the first authors to announce

[3] See, for example, Artzrouni and Komlos (1985).

[4] Hardin (1999).

[5] Or, in Wicksteeds words: "For it is one thing to be practically familiar with a principle and to assume it in simple cases as a matter of course, and it is another thing to grasp it so consciously and so firmly as never to lose hold of it or admit anything inconsistent with it, however remote from familiar experience and however complicated and abstract may be the regions of enquiry in which we need it as our clue." Wicksteed (1894), p. 8.

a “law of diminishing returns to labor”, according to which some constant production factor (e.g. capital, land) would limit the rise of production induced by an increase in the labor force. A few years later, A. Smith[6] (1776) partly revised this physiocratic view in the light of the English Industrial Revolution by stating that higher population density and urbanization would cause a greater variety of professions, raising the degree of specialization. If increasingly specialized individuals would reasonably engage in trade, the “division of labor” between these subjects would be enhanced, raising production more than proportionally. Another twenty years later however, the idea that the wealth of nations was based on population growth was struck again when it had become clear that in spite of great technological advances resulting from the division of labor, the British population explosion had effectively pushed down real wages and GDP per capita. Malthus (1798) proposed the “principle of population”, by stating that population had the inherent tendency to inevitably outgrow production. Another five years later however, as we will show in Part II of this work, Malthus (1803) provided the “great preventive check” as apparently constituting the only justifiable remedy for economies facing excessive population growth and by which individuals were generally susceptible to birth control. Since then, as predicted by the first professor of political economy, fertility abated and productivity increased.

3.1.3 Endogeneity of the Malthusian trap

To extend our theoretical understanding of stagnation in productivity from a rather ad hoc relationship between production growth and population growth toward the classical, more sophisticated endogenous growth theory, we will distinguish between a classical theory of production and a classical theory of population. Accordingly, we will explore the effect of population growth on productivity growth ($\frac{\partial g_y}{\partial g_N}$) as well as the effect of productivity growth on population growth ($\frac{\partial g_N}{\partial g_y}$) separately. The underlying principles guiding these effects must be verifiable and universal, even though they may outweigh each other to the extent that some principle may not be perceptible at a given point in time. Consequently, in the following sections much emphasis is put on exhaustive definitions and logical deductions. Once we have uncovered the main principles of the theory of production as well as of the theory of population, we may unify them in one coherent endogenous growth framework to simulate an era of stagnation. We will now start with the classical theory of production.

[6] Adam Smith (1723–1790), British professor of moral philosophy at the University of Glasgow, one of the founders of classical economics/political economy.

3.2 The Classical Theory of Production

> The annual labour of every nation is the fund which originally supplies it with all the necessaries and conveniences of life which it annually consumes [...][7]

3.2.1 A labor theory of production

3.2.1.1 Introduction

Firstly, to frame a long–run demographic theory of production as outlined in chapter one, growth in production must be *exhaustively* defined. An important benefit of using an exhaustive definition of production is that it allows us to structure our economic thinking by "keeping everything in mind at the same time." To account for an exhaustive theory of production, we will employ the (static) neoclassical macroeconomic aggregate production function approach that incorporates a relationship between production and every factor of production. Instead of explicitly accounting for "technology", all other production factors apart from population will be collected in the variable "generalized capital." This section will show that this traditional production function approach is largely in line with classical theory. Secondly, while production may be defined as "all value produced within a given year", it is essential to model population as *only source* of production. Only if these two preconditions are fulfilled for a theory of production and analogously for a theory of population, the latter two might be synthesized into an endogenous theory of growth that can sufficiently explain the process of economic development.

In this first section, building on the Smithian assumption of an efficient division of labor, the (static) production function approach by P. Wicksteed[8] (1984) is extended toward the (dynamic) Solow growth model, where (exogenous) labor is assumed to affect all other production factors in addition to production itself. In the second section it will be clarified that what was commonly meant by the variable "labor" should be substituted by the variable "unskilled labor," which is well approximated by population size. On this basis, a new model of population as *source of all value* is presented, where regular innovations are supposed to be completely embodied by population growth through a division of labor and subsequent specialization and a structural equation, defining the simultaneous operation of two universal economic principles, is derived. This structural equation will later be used for the simulation and estimation of our model.

3.2.1.2 Static Theory

The one decisive regular cause by which population growth is classically assumed to enhance production is the *division of labor*. Smith (1776) suggested that, in an environment favoring the security of property and income, the inherent tendency to exchange their products would result in a division of labor between individuals.[9] What Smith meant by "division of labor" can be

[7] Smith (1776), Introduction.
[8] Philip Henry Wicksteed (1844-1927), British Unitarian minister, one of the founders of neoclassical economics.
[9] Throughout this paper we will presume the existence of such institutional conditions. A more abstract point of view would even suggest relaxing this presumption by stating that efficient institutions are one result of such a perfect division of labor.

understood as the efficient cooperation of all productive individuals in the economy to maximize total production. This efficient level of cooperation would be achieved if all production processes were perfectly subdivided between those individuals. Such a perfect division implied that every new individual entering the economy would tend to induce a new subdivision of production into smaller, more easily conductible, efficient production processes and therefore raise production.

Nonetheless, in spite of the generally observed tendency toward an efficient division of labor in a free market economy, the classics had already noted large regional differences in individual productivity. Obviously, these differences were owed to the relative abundance or scarcity of some other production factor. N. W. Senior[10] (1836) extended the doctrine of labor division by stating that what Smith had really meant was the efficient subdivision of production processes through an efficient combination of *several* production factors.[11] He argued that this concept ought in fact be termed the "division of production" instead of the "division of labor" and could be formulated as a relationship between production and an efficient use of all these production factors – a production function.[12]

Perpetuating this production function approach, neoclassical economists argued at the end of the nineteenth century that, since production factors are in reality dynamically interconnected, a separate analysis of the effect of every single factor required the consideration of an abstract static state of an economy where production factors were assumed to be independent. To J. B. Clark[13] (1899), it seemed obvious that the economist had to start with the easier task of modeling a static production function first, where all except the production factor of interest were held constant (*ceteris paribus*) such that no causal relationships between production factors interfered.[14] Roughly at the same time, the static model was mathematically advanced by Wicksteed (1894), adding to the above considerations the perhaps most powerful proposition for employing a valid aggregate production function: *the replication argument.* It states that under static conditions, a replication of an exhaustive list of production factors must universally generate a replication of production. Correspondingly, an aggregate production function is to be defined as a static production function fulfilling the replication argument, which was later formulated as the doctrine of constant returns to scale.[15]

> Now it must of course be admitted that if the physical conditions under which a certain amount of wheat, or anything else, is produced were exactly repeated the result would be exactly repeated also, and a proportional increase of the one would yield a proportional increase of the other. The crude division of the factors of production into land, capital and

[10] Nassau William Senior (1790–1864), British lawyer, professor of political economy at the University of Oxford, member of Royal Commissions in 1832, 1837 and 1861.

[11] A production factor being defined as an input resource that positively contributes to production.

[12] "[...] division of production would have been a more convenient expression than division of labour; but Adam Smith's authority has given such currency to the term division of labour, that we shall continue to employ it, using it, however, in the extended sense in which it appears to have been used by Adam Smith." Senior (1836), p. 159.

[13] John Bates Clark (1847-1938), US–American professor of economics at Columbia University, founder of the American Economic Association.

[14] "Why, then, do we wish to know the laws of an imaginary static state? Because the forces that act in such a state continue to act in a dynamic one. [...] In dealing with the complex problems of an advancing economy, the key of success is the separate study of the static forces that constantly act within it." Clark (1899), p. 60.

[15] See Hicks (1936).

> labour must indeed be abandoned [...]. We must regard every kind and quality of labour that can be distinguished from other kinds and qualities as a separate factor; and in the same way every kind of land will be taken as a separate factor. [...] Each of these may be scheduled in its own unit, and when this has been done the enumeration of the factors of production may be regarded as complete. On this understanding it is of course obvious, that a proportional increase of all the factors of production, will secure a proportional increase of the product.[16]

Notwithstanding the requirement of an exhaustive list of factors, most classical economists seem to have agreed on the usage of merely two factors required for production Y: labor L and capital K.[17] Wicksteed declared the just use of this simplification as long as capital was viewed to serve as an approximate residual "catch–all–variable" to incorporate an exhaustive list of all hitherto omitted production factors required for total production, measured in a complex unit, for example in exchange value.[18] As a result, capital could be defined as all *things of value required for production*, whereas the explicit use of other production factors would have unnecessarily complicated the theoretical and empirical analysis – a simplification that was later used by Keynes, Robinson[19], Solow[20] and many others. Obviously, when following this definition, a potential factor specifically accounting for "technology"[21] would become obsolete.

Alongside this generalized capital, labor remained the main production factor of interest as long as labor productivity $\frac{Y}{L}$ seemed to best approximate individual productivity, i.e. production per capita. As J. R. McCulloch[22] (1863) concluded,

> [he]re, then, is the simple and decisive test by which we are to judge of the expediency of all measures affecting the wealth of the country, and of the value of all innovations. If they make labour more productive, [...] they must be advantageous; [...] Considered in this point of view, that great branch of the science which treats of the production of wealth will be found to be abundantly simple, and easily understood.[23]

To this end, Wicksteed suggested that labor had to be isolated from an infinite number of production factors:

> What we really want is to separate out labour and dose it with land–plus–capital, if possible to satiety.[24] [. . .] It is perfectly legitimate to start with a unit of [labour], assume that

16 Wicksteed (1894), p. 33.

17 Often, the additional factor land has been added: "[. . .] it has been usual to take each of the great factors of production such as Land, Capital and Labour, severally, to enquire into the special circumstances under which that factor co–operates in production [. . .]." Wicksteed (1894), p. 7. Thereafter, a production function with constant returns to scale is mathematically defined as follows: $F(K, L) = Y \Leftrightarrow F(\lambda K, \lambda L) = \lambda Y$.

18 "All the constituents of this generalised 'capital' are regarded as reduced to their expression in money." Wicksteed (1894), p. 13.

19 "The capital in existence at any moment may be treated simply as 'part of the environment on which labour works.'" Keynes (1936) in Robinson (1954), p. 214.

20 "Were the data available, it would be better to apply the analysis to some precisely defined production function with many precisely defined inputs. One can at least hope that the aggregate analysis gives some notion of the way a detailed analysis would lead." Solow (1957), p. 312, footnote.

21 As is suggested by the use of the production function $Y = F(A, K, L)$, which is for example done by Solow (1957) or Romer (1986).

22 John Ramsay McCulloch (1779–1864), British professor of political economy at London University, comptroller of Her Majesty's Stationary Office.

23 McCulloch (1863), part I, chapter I, section II.

24 Wicksteed (1894), p. 14, footnote.

> the command of the other factors of production is so exercised as to secure the maximum productive result, and then treat the product as a function of [labour] and pounds sterling. [...] and we may, if we choose, select any one factor to measure in its proper unit while measuring all the rest in a common unit.[25]

Furthermore, Wicksteed concluded that such an aggregate production function implied diminishing returns to each production factor in the fashion forwarded by Turgot (1767) and von Thuenen (1842), i.e. that an incremental static use of any separate factor would yield an increasingly diminishing marginal product as well as diminishing productivity of that factor.[26] Hence, the idea of diminishing returns became a universal law for the accumulation of any production factor, which is one of the conclusive statements of the static theory of production.[27] Cobb and Douglas (1928) built on Wicksteed's approach and suggested a specific form of an aggregate production function that would account for the above conditions.[28] Their aggregate production function $Y = F(K, L) = K^{\alpha}L^{1-\alpha}$ with $0 < \alpha < 1$, where α reflects the constant production elasticity of capital, is still a commonly taught instrument of the neoclassical growth school, although it does not necessarily display the "true" form of the aggregate production function.

In summary, we may conclude that Smith's classical concept of labor division has survived the marginal revolution in the form of a neoclassical aggregate production function centering on the production factor labor. Consequently, the static theory of production supports the idea that each additional amount of labor entering an efficient division of labor causes production to rise and labor productivity to shrink due to diminishing returns. However, it does not tell us anything about the interdependencies of the variables over time.[29]

3.2.1.3 Dynamic Theory

By merely assuming that all production factors are efficiently employed, thus far nothing has been said about the potential dynamic effects, or gains, derived from a division of labor. On this account, Smith emphasized that

> [T]he greatest improvement in the productive powers of labour, and the greater part of the skill, dexterity, and judgment with which it is anywhere directed, or applied, seem to have been the *effects* of the division of labour.[30]

---

25 Wicksteed (1894), p. 39.

26 "Then if [labour] remains constant and capital–plus–[land] increases, we shall have increasing returns per unit of [labour] and decreasing returns per unit of capital. But if capital is constant and [labour] increases, we shall have increasing returns per unit of the former and decreasing returns per unit of the latter." Wicksteed (1894), p. 14, footnote

27 See also Humphrey (1997).

28 "The theory referred to (due to J.B. Clark, Wicksteed et al.) states that production, labor and capital are so related that [...] production is a first degree homogeneous function of labor and capital." Cobb and Douglas (1928), p. 151.

29 "The dimension of time enters negatively into all the quantities we are discussing. 'Land' is use of land per unit of time. Labour is hours of work per unit of time, etc. But the universality of this condition enables us to dispense with any special consideration of it." Wicksteed (1894), p. 20, footnote.

30 Smith (1776), book I, chapter I. Italics by the author.

The development of the subdivision of production processes can be illustrated using a simple example. One might conceive of a potato farmer owning a hectare of land. Having spent all his life in farming, he gradually developed a wooden plow to simplify his efforts. If one were to duplicate this one–man–economy, constant returns to scale would require all production factors, i.e. the farmer, his accumulated skills in building the plow and tilling the soil, one hectare of land as well as the plow and the remaining quality of the land. The result would be two economies with two economic agents and with output per person staying unchanged. If some factor (e.g. land) would remain constant, diminishing returns would even decrease production per capita. However, as soon as the dynamic factor time is introduced, the farmers will, given this option, start to communicate on their methods of production and will very soon realize the merits of cooperation and specialization. Once they decided to specialize into two different tasks, say the one in building plows and the other one in tilling the soil, the above duplication of production factors yields "after some time" a more than proportional total increase in output that can make up for possible losses due to diminishing returns.

Extending the static theory of production toward a dynamic theory of growth, Solow (1956) and Swan (1956) integrated the Cobb–Douglas production function into Harrod's (1939) and Domar's (1946) concepts of intertemporal capital accumulation by using as central dynamic equation $K_{t+1} = sY_t + (1-\delta)K_t$, or in units per labor

$$\frac{K_{t+1}}{L} = s\left(\frac{K_t}{L}\right)^{\alpha} + (1-\delta)\,\frac{K_t}{L} \tag{3.2}$$

with time index t for the corresponding year, annual savings rate s and annual capital depreciation rate δ. Considering the direction of causality between the production factors, this framework assumes that the amount of labor is exogenously supplied, while the amount of capital is allowed to adapt over time. The level of labor is therefore assumed to be unaffected by capital changes, whereas changes in the amount of labor may generally cause changes in the amount of capital. Again, the rationale behind the exogenous use of labor as compared to capital can be traced back to classical economics and in particular to Locke (1689) and McCulloch (1863), who considered labor as the only source of value, without which capital would be not worth anything:

> 'Tis labour, then which puts the greatest part of value upon land, without which it would scarcely be worth of any thing. Tis to that we owe the greatest part of all its useful products;' [...] Locke has here all but established the fundamental principle on which the science [of economics] rests. Had he carried his analysis a little farther, he could not have failed to perceive that neither water, leaves, skins, nor any one of the spontaneous productions of nature, has any value, except what it derives from the labour required for its appropriation. The utility of such products makes them be demanded; but it does not give them value. This is a quality which can be communicated only through the agency of voluntary labour of some sort or other. [...] It is to labour, therefore, and to it only, that man owes every thing possessed of value.[31]

[31] Locke (1689) "Of Civil Government" book ii §§ 42, 43 in McCulloch (1863), part I, chapter I, section II.

On these grounds, dynamic changes in labor productivity can be modeled as a response to an exogenous labor shock as follows. As a starting point, neoclassical economists reasonably assume a static equilibrium in which capital accumulation is equal to zero and capital depreciation equals saving, i.e. for $K_{t+1} = K_t = K$ we have

$$\delta \frac{K}{L} = s \left(\frac{K}{L} \right)^{\alpha}. \tag{3.3}$$

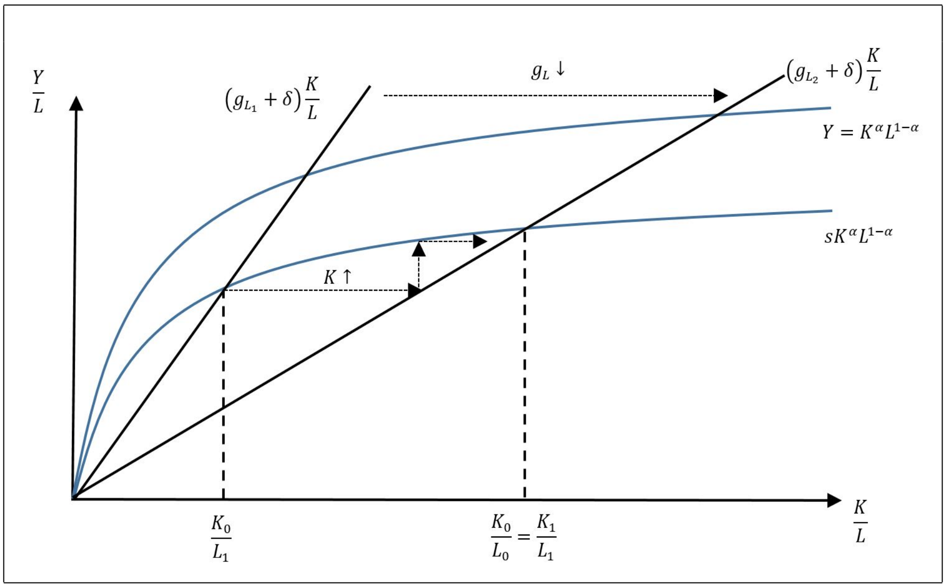
The resulting "steady state" equilibrium of capital per labor $\left(\frac{K}{L}\right)^* = \frac{K_0}{L_0}$ is marked on the x–axis of Figure 3.1 (ignoring g_{L_2} at this instance). In this situation, a positive labor shock would statically reduce capital per labor toward $\frac{K_0}{L_1}$, where savings are higher than capital depreciation. Subsequently, additional capital will be accumulated and capital per labor eventually reconverges to its original steady state such that $\frac{K_0}{L_0} = \frac{K_1}{L_1}$ with $K_1 > K_0$ and $L_1 > L_0$. As a result, although diminishing returns have reduced labor productivity in the short run, a growing labor force seems to be capable of accumulating and maintaining a larger amount of capital in the long run, reflecting the abstract "gains" from a division of labor, which will be examined in more detail in the next section. The same mechanism applies reversely if labor shrinks. In that case, relative labor scarcity increases labor productivity in the short run without being able to maintain the old amount of capital in the long run. As a general result, it might be deduced that every change in the variable labor is in the long run followed by a proportional change in the variable capital such that we may write in terms of growth rates

$$g_Y = \alpha g_K + (1 - \alpha) g_L = g_L, \tag{3.4}$$

where g_Y, g_K and g_L denote the growth rates of production, capital and labor respectively.[32] Consequently, labor productivity $\frac{Y}{L}$ would in this framework ultimately remain constant after a labor shock, since we have $g_{Y/L} = g_Y - g_L = 0$.

[32] As Solow emphasized, "[T]he common rate of growth is just the exogenously given growth rate of the labor force." Solow (2001), p. 357.

Figure 3.1: A labor growth slowdown in the Solow model.



It is often argued that the conventional Solow model is incomplete as it can not account for long–run changes in labor productivity.[33] This claim often overlooks the effect stemming from a potential change in labor growth. Supposing that labor would increase every period at the same constant rate $g_L = \frac{L_{t+1}}{L_t} - 1$ yields the following modified dynamic law and steady–state value for labor productivity:

$$(1+g_L)\frac{K_{t+1}}{L_{t+1}} = (1-\delta)\frac{K_t}{L_t} + s\left(\frac{K_t}{L_t}\right)^{\alpha} \Leftrightarrow \left(\frac{Y}{L}\right)^{*} = \left(\frac{s}{\delta+g_L}\right)^{\frac{\alpha}{1-\alpha}}. \tag{3.5}$$

Since, as can be seen from the right equation of 3.5, higher labor growth reduces the steady–state value of labor productivity, we find that an exogenous decrease in labor growth from g_{L_1} toward g_{L_2} (see Figure 3.1) is well–qualified for causing labor productivity to increase.

3.2.2 A demographic theory of production

3.2.2.1 Static Theory: The Principle of Diminishing Returns

To test the validity of the above neoclassical model, empirical long–run estimates for production Y_t are, as we have seen in the last chapter, generally readily obtainable. However, since generalized capital K_t can only be measured as a residual, we require empirical values for labor L_t as well as those for the production elasticity of labor $(1-\alpha)$. To this end, the crucial questions arise concerning how labor is supposed to be defined in theory and in what units it thus ought to be measured empirically.

[33] As a direct consequence, the additional factor "technology" (Solow (1956)) or "measure of our ignorance" (Abramovitz (1956)) was introduced , embodied in the so–called "Solow–Residual."

Theoretically, we will again follow the classical view of Senior (1836), who defined labor as "the voluntary exertion of bodily or mental faculties for the purpose of production."[34] In addition, McCulloch seems to have sufficiently differentiated as to how far labor has to be viewed as part of the production process:

> So long as an individual employs himself in any way not detrimental to others, and accomplishes the object he has in view, his labour is obviously productive; while, if he do not accomplish it, or obtain some sort of equivalent advantage from the exertion of his labour, it is as obviously unproductive. This definition seems sufficiently clear, and leads to no perplexities; [...] it is not possible to adopt any other without being involved in endless difficulties and contradictions.[35]

Such a definition comprises the quality and the quantity of labor, or to use a slightly different modern wording, skilled as well as unskilled labor.

Empirically, the first assessment of economic growth based on the aggregate production function used in the former section was conducted by Cobb and Douglas (1928). Problematically, they used Wicksteed's production exhaustion theorem[36] to interpret the empirical share of labor income on total income as production elasticity of *labor* $(1-\alpha)$, whereas they measured the production factor L in units of *laborers* with the following reservation:

> Such an index $[L]$ of course makes no allowance for possible changes in the quality of the laborers [...]. When they can be measured, then they should be included.[37]

Notwithstanding this qualification, current empirical and theoretical studies still seem to erroneously follow Cobb and Douglas' provisional model and continue to confuse the production elasticity of labor with that of the number of laborers without considering the quality of the laborers.[38] To make up for this defect we may first attempt to assess whether a labor measurement exists that can account for the quality as well as the quantity of labor.[39] However, once finding that it is still not possible to measure a unit of labor quality without making fantastic assumptions, it will – in pursuing the population question and in explaining economic development – appear much more promising to focus on a theoretical conception that separates out the quantity of labor. We will therefore now discuss the necessary adjustments for a neoclassical model based on unskilled labor as central variable.

34 Senior (1836), p. 152.

35 McCulloch (1863), part I, chapter I, section II.

36 Since $F(K,L)$ is homogeneous of degree one, the Euler theorem can be applied as follows, where r represents the marginal product of capital and w the marginal product of labor:

$$Y = F(K,L) = K^{\alpha}L^{1-\alpha} = \frac{\partial Y}{\partial K}K + \frac{\partial Y}{\partial L}L = \alpha Y + (1-\alpha)Y = rK + wL \Leftrightarrow 1-\alpha = \frac{wL}{Y}.$$

"[...] under ordinary conditions of competitive industry, it is sensibly or approximately true that if every factor of production draws a remuneration determined by its marginal efficiency or significance, the whole product will be exactly distributed." Wicksteed (1894), p. 38.

37 Cobb and Douglas (1928), p. 149.

38 Barro and Sala–i–Martin (2003), for example, use units of *labor* instead of units of *laborers, workers* or *population* in order to calculate output per worker and output per capita, pp. 27–28.

39 Unfortunately, while the Cambridge capital controversy has questioned the correct measurement of the production factor capital, no such debate can be found on the empirical use of the production factor labor.

In contrast to Cobb and Douglas, who employed the number of laborers to measure the overall amount of unskilled labor, this work intends to use the size of the population since our analysis of economic development centers around the variable production per capita, i.e. population productivity, instead of productivity per laborer.[40] From a theoretical point of view this seems perfectly valid, since unskilled labor moves – although with a maturity lag – almost proportionally with population. Empirically, this substitution greatly simplifies and improves the subsequent economic and econometric analysis.

To exhaustively subdivide labor into two independent components, we will then refer to the production factor displaying the quantity of labor as population N, while terming the remaining factor human capital H, incorporating the residual quality of labor. The latter is supposed to comprise every acquired productive skill in addition to unskilled labor. Statically, like every other production factor, human capital and population must necessarily exhibit diminishing returns – a notion we will use in the following to express our first universal principle, "the principle of diminishing returns to population" (in the following "PoDR"). Following Mankiw et al.'s (1992) extension of the Solow model, this subdivision suggests the use of the following production function:[41]

$$Y = K^{\alpha} L^{1-\alpha} = K^{\alpha} H^{\beta} N^{1-\alpha-\beta} \; with \; 0 < \alpha, \beta, (1-\alpha-\beta) < 1.$$

[40] Firstly, we define an unskilled individual i as an individual who has just entered the labor market and has therefore become productive for the first time. We may then define a unit of unskilled labor l_u as the amount of labor provided by such an unskilled individual. Eventually, the aggregate amount of unskilled labor L_u is defined as one unit of unskilled labor multiplied by the number I of all productive individuals, i.e. $L_u = i \cdot l_u \cdot I$. Since we are interested in the annual amount of unskilled labor of the *average* individual, i and l_u can be standardized to 1.

[41] This paper makes two important deviations from the Mankiw et al. (1992) model: Population is used instead of labor and total factor productivity is assumed to be non–existent, i.e. $A = 1$, following the Wicksteed (1894) approach.

3.2.2.2 Dynamic Theory: The Principle of Labor Division

Given our new static division of production, we will now again inquire into the relationships between the production factors and account for the currently popular concept "technology". In the former subsection, we assumed that the decisive causal relation runs from labor, the source of all value, to capital accumulation. We will see that this causal effect can be renewed inasmuch as population growth may be considered as the source of all human capital accumulation as well as physical capital accumulation – a result that can be derived if we take a more detailed, microeconomic look at Smith's dynamic effects from the division of labor. According to Smith, all men are born equal and every worker acquires in the same way both productive skills and productive capital over his lifetime to optimize individual production. Senior remarked about Smith's central idea:

> The advantages derived from the division of labour are attributed by Smith to three different circumstances. 'First, to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labour, and enable one man to do the work of many.'[42]

These effects are all reducible to gains from specialization and can account for an exhaustive dynamic theory of growth. Firstly, "the saving of time" is owed to specialization across a given territory. Obviously, as soon as new individuals are added to the division of labor, the economy becomes more densely populated and the efficient geographical distribution reduces any kind of transport costs between production processes. This advantage simply reflects the static use of the factor N including diminishing returns through the relative abundance of this factor. Secondly, "the increase in dexterity" is owed to specialization over time. If the same production process is performed frequently on tighter geographical constraints, individuals will successively tend to improve their productive skills by way of learning and subsequently use their experience to become a specialist in their field.[43] Since skilled labor H can only be accumulated by repeatedly employing unskilled labor N, population growth can be rightfully viewed as the only source of human capital accumulation. Thirdly, "the invention of machines" refers to a regular tendency toward automation of specialized processes. Whenever production processes have been subdivided into such small steps that their repetition can be easily conducted through some non–human agency, capital K tends to be substituted for labor H and N. As a result, the accumulation of any production factor is rooted in the use of unskilled labor, approximated by population. Every increase in population will eventually raise the amount of the other production factors, raising in turn production Y and population productivity $\frac{Y}{N}$. Accordingly, we may state that population, "the starting and ending point of all economic activity"[44], is really the source of all value and the most regular trigger of economic growth – a notion we will refer to in the following as "the principle of labor division" (in the following "PoLD"). Consequently, since the demographic

[42] Smith (1776) in Senior (1836), p. 159.

[43] "Men are much more likely to discover easier and readier methods of attaining any object, when the whole attention of their minds is directed towards that single object." Smith (1776), book I, chapter I.

[44] Bairoch (1988), p. 127.

theory is the only view capable of explaining long–run growth in production *exhaustively*, it is the only valid view of a unified growth theory. "Hence the peopling process is essential"[45] and we shall begin every inquiry on economic growth by examining the effects stemming from any foregoing population changes. As Young (1928) put it,

> Senior's positive doctrine is well known, and there were others who made note of the circumstance that with the growth of population and of markets, new opportunities for the division of labour appear and new advantages attach to it. In this way, and in this way only, were the generally commonplace things which they said about 'improvements' [. . .].[46]

Based on Smith's theoretical considerations of the gains from a division of labor, we can derive additional static and dynamic interpretations of the neoclassical growth model. Since capital K and human capital H are in the same way frequently and proportionally accumulated after new productive individuals have entered the economy, we should reasonably assume that human capital is subject to the same law of accumulation as capital ($\delta_H = \delta_K \equiv \delta,\ s_H = s_K \equiv s$) and can be measured – like any other production factor – in the same complex unit. Thus, we can make use of a model, where population N is separated out of the inifinite number of production factors and where human capital and physical capital are aggregated into one and the same complex production factor *generalized capital 2.0*, or *broad capital* C, as follows.[47]

$$Y = K^{\alpha} H^{\beta} N^{1-\alpha-\beta} \equiv C^{\gamma} N^{1-\gamma} \ \textit{with}\ 0 < \gamma < 1$$

In analogy to equation 3.5, we write our new dynamic law of broad capital accumulation as

$$(1 + g_N) \frac{C_{t+1}}{N_{t+1}} = s \frac{Y_t}{N_t} + (1 - \delta) \frac{C_t}{N_t}$$

With regard to the causal relation between production factors, we conclude again that, since exogenous changes in population are the source of all value, broad capital would in the long run react proportionally to population. Consequently, population growth would once more cause capital growth and economic growth without, however, raising production per capita. The corresponding stable steady state with $\frac{C_{t+1}}{N_{t+1}} = \frac{C_t}{N_t} = \frac{C}{N}$ is given by

$$\left(\frac{Y}{N}\right)^{*} = y^{*} = \left(\frac{s}{\delta + g_N}\right)^{\frac{\gamma}{1-\gamma}} = \left(\frac{s}{\delta + b - d}\right)^{\frac{\gamma}{1-\gamma}} = \left(\frac{s}{b}\right)^{\frac{\gamma}{1-\gamma}}$$

where we have defined the constant (natural) population growth rate as $g_N = b - d =$ *(crude) birth rate – (crude) death rate*[48] and made the simplifying assumption that over the long–run

[45] Lange (2012), p. 21.

[46] Young (1928), p. 529.

[47] Intuitively, the argumentation follows Smith: "The improved dexterity of a workman may be considered in the same light as a machine or instrument of trade which facilitates and abridges labour, and which, though it costs a certain expense, repays that expense with a profit." Smith (1776), book I, chapter I.

[48] The birth rate and the death rate, both generally denoted as vital rates, are defined as the ratios of births and deaths respectively to the total population size in a given year. With the natural change in population size being given by $\Delta N = \textit{Births} - \textit{Deaths} = B - D$, the natural population growth rate can be computed by $g_N = \Delta N/N = (B - D)/N = b - d$.

$\delta = d$, since the skills of a population as well as the (unskilled) population itself depreciate with the death rate ($\delta = \delta_H = \delta_N = d$).[49] Allowing for a varying birth rate, the productivity ratio of two subsequent steady states providing the following "unit–free" measurement of the growth factor would also depend on the birth rate:[50]

$$\frac{y_t^*}{y_{t-j}^*} = \left(\frac{b_{t-j}}{b_t}\right)^{\frac{\gamma}{1-\gamma}} \quad \textit{with } 0 < \gamma < 1. \tag{3.6}$$

In summary, this intertemporal representation of productivity allows us to distinguish between the essentially conflicting effects of population growth in the theory of production. While the numerator of the right hand side of equation 3.6 has a delayed positive effect on productivity growth, representing the gains from labor division, the denominator affects productivity immediately negatively, representing the losses from diminishing returns. Ultimately, the population question boils down to these two opposing forces describing the essential conflict by which we have to judge population growth economically.

> The 'population' question [...] will then be found ultimately to turn on a balance between the significance to each man of other free men regarded as appliances and the significance to him of the space those other men occupy. Is their room [effects from diminishing returns] or their company [effects from labor division] the more important?[51]

From equation 3.6 it also becomes apparent that the steady–state growth rate of productivity is determined neither by the level of population nor by population growth itself. Instead it is governed by the (inverted) *growth of population growth*. As is illustrated in Figure 3.2, whenever the birth rate increases, the losses from diminishing returns tend to outweigh the gains from labor division and to create a situation of economic regress. This subject of overpopulation due to diminishing returns seems equally obvious as scientifically neglected or avoided.

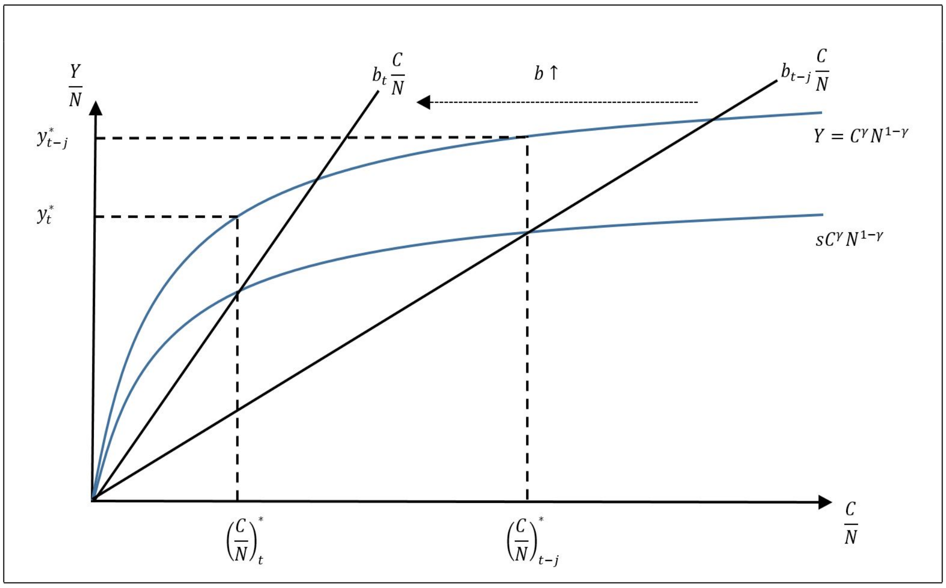
---

49 Obviously, if the death rate is zero (and the birth rate positive), population grows infinitely together with human capital and physical capital. However, if population is heavily written off due to a high death rate, H is increasingly lost and K cannot be maintained, eventually depreciating. Commonly, depreciation rate as well as death rate both lie within the range $[0.01, 0.04]$ (see also Mankiw et al. (1992), p. 410). These observations lead to the conclusion that death rate and depreciation rate move proportionally.

50 The author is well aware of the vast literature on endogenized savings. However, this issue should be separated from attempting to answer the population question. Consequently, we will hold aggregate savings constant throughout this work.

51 Wicksteed (1894), p. 22, footnote.

Figure 3.2: A population growth slowdown in the Solow model.



3.2.3 Recapitulation: A demographic theory of production based on the division of labor and diminishing returns

In this brief exercise on the theory of production, the conventional Solow model has been modified by renewing classical assumptions. Firstly, the production factor labor has been replaced with the production factor population. Secondly, the model expressly excludes exogenous "technology". In particular, instead of modeling "technology" endogenously, the following assumptions have been used to account for economic innovation. As the Cobb–Douglas production function is theoretically based on Smith's assumption of an efficient division of labor, the accumulation of production factors is theoretically based on Smith's *gains* from the division of labor, which are made up of specialization across space and time as well as automation of repetitive laboring processes. As a side note, this means that every improvement that is commonly very loosely termed "measure of our ignorance" should be included in the production factor "broad capital." Thirdly, as we have seen, it is a well–established fact in neoclassical economic theory that increasing the amount of population tends to increase overall production. Nevertheless, by holding the stock of all other production factors constant, a growing population is generally acknowledged to yield diminishing returns, i.e. to decrease production per capita. The mathematical formulation of this effect is displayed by the use of the static neoclassical production function developed by Wicksteed (1894) and popularized by Cobb and Douglas (1928). Nonetheless, fourthly, since those diminishing returns are successively compensated by an increasing division of labor, Solow (1956) answered the population question by stating that every increase in population has a neutral impact on productivity over the long run. To allow for a clear empirical distinction between the static effect of diminishing returns (PoDR) and the dynamic effect of labor division (PoLD) on production, the analysis will subsequently be greatly simplified.

Firstly, the PoDR is assumed to provide a negative "static" causal effect running from population to productivity and was obviously the decisive effect omitted in the Kremerian framework. This negative effect will be modeled using a contemporaneous relationship between GDP per capita y and population growth g_N, which is reduced to changes in the birth rate b. The resulting causal effect by which a newborn individual will by definition instantly affect GDP per capita might be written as $\frac{\partial g_{y_t}}{\partial b_t} < 0$.

Secondly, the PoLD relates the production factor population positively to its level of production, yielding the gains derived from the division of labor. The effect stemming from the PoLD can be interpreted to roughly correspond to the Kremerian (or Boserupian (1965)) idea by which a larger population raises the chance of discovering more productive innovations, embodying the conventionally used concept "technology". Nonetheless, the Smithian principle is less owed to probability, but the logical consequence of a more sophisticated process of specialization. However, an increase in population will not yield benefits from the division of labor contemporaneously, but rather lagged. With respect to a newborn individual, the minimum delay to account for a positive increase in production as a response to an increase in population is given by the time span reserved for a basic education, enabling the succeeding generation to participate in the labor market, i.e. to be "productive". For simplicity and as it is sufficient to illustrate the role played by the PoLD in the classical framework, the effect of *only one* birth cohort per generation will subsequently be employed in our model. Furthermore, we also make the assumption that all gains from an efficient labor division are realized within *the first year* in which an individual enters the labor market, determining the growth rate in a given year by writing $\frac{\partial g_{y_t}}{\partial b_{t-g}} > 0$, with g accounting for the generational lag. Since the overall positive effect of population growth on production operates in reality through the gradual accumulation of capital over a number of years, it is not easy to simulate. We will return to this effect more precisely when evaluating the model empirically in the last chapter. At present, to provide a simple linear relationship of the theory of production, equation 3.6 is reduced to the following representation:

$$g_{y_t} = \overbrace{ln\ b_{t-g}}^{PoLD} - \overbrace{ln\ b_t}^{PoDR} \tag{3.7}$$

Verbally, the principles might be formulated as follows. Firstly, that at the very moment of entering into the economy, every additional individual will statically lower production per capita (PoDR). Secondly, that with a delay of approximately one generation, production per capita responds positively, proportionally and indefinitely to an increase in population under the condition that the additional part of the population participates in the division of labor of the economy (PoLD).

3.3 The Classical Theory of Population Part I: The Principle of Population

> Population, we must allow, is one of a numerous class of subjects which serve to remind us that man has frequently left him only a choice of evils. Whatever may have been its design, suffering has entered into the plan of Providence. [...] It is of no use to rebel against this order of things; for it envelopes us; it is the atmosphere in which we live and breathe; and it is with [the] alternative of restriction and prevention before us, which we cannot get rid of, and cannot lose sight of that we proceed, with Malthus, to enter upon the problem of population.[52]

In section 3.2, we have assumed exogeneity of the population growth rate and endogeneity of the growth rate of production. Having thus modeled the impact of population on productivity, the third classical principle defined in this section determines the impact of productivity on population. First and foremost, every classical theory of population must be based on the principle of population. In contrast to the PoDR and the PoLD, this third principle is much less utilized in neoclassical models. Thence, this section provides a new introductory account of the principle of population. As a by–product, prevalent misconceptions will be clarified. Since some confusion seems to exist with regard to the terms "principle of population" and "Malthusian trap", this work intends to settle the distinction. To adhere as closely as possible to Malthus' own, undistorted thoughts, deductions will often be sustained by quoting Malthus himself. Also, as it is the author's conviction that the significance of the principle is most thoroughly grasped by evaluating the intellectual impact it exerted on some of the most celebrated contemporary scientists, these corresponding authorities, providing first-hand evidence, will be frequently cited as well.

3.3.1 The power of population

Malthus' (1798) first important presumption on the theory of population was to state that every population possessed the power to grow exponentially, or as the classical authors used to call it, "in a geometric ratio". Although it became controversially debated during the first years after its appearance, the presumption was soon well-received among the profession of political economists. By the year 1836, Senior had outlined the classical theory of population, beginning with the assertion that

> it is now generally admitted, indeed it is strange that it should ever have required to be pointed out, that every species of plant or animal which is capable of increase, either by generation or by seed, must be capable of a constantly increasing increase.[53]

Likewise, Mill (1848), referring in turn to Senior, granted the power of population an important role in his "Principles of Political Economy":

---

[52] Bastiat (1860), p. 400.
[53] Senior (1836), p. 141.

> To this property of organized beings, the human species forms no exception. Its power of increase is indefinite, and the actual multiplication would be extraordinarily rapid, if the power were exercised to the utmost.[54]

To provide an illustration, the following calculation will demonstrate the power of (hypothetically) unregulated exponential population growth. It has been estimated that the global human population of the year 1804 amounted to about one billion people.[55] If the maximum life expectancy was assumed to be eighty years, which is certainly under the mark, and with maximum fertility having been calculated at about 16.7 children per woman[56], these values imply, given a stationary[57] population, a birth rate of 10.43% and a death rate of 1.25%.[58] The maximum natural[59] growth rate of population can therefore be computed by

$$g_N = \Delta N/N = (B - D)/N = b - d \tag{3.8}$$

to be 9.19%. Thus, if the power of population had operated unrestrictedly since the year 1804, the correspondingly projected population size would in the year 2017 have amounted to 135,155,105 billion inhabitants, i.e. the average person would have produced over 135 million descendants after 213 years.[60] For other species, the case can be even more strikingly portrayed. H. Spencer[61] (1852) reported in the context of Malthus' theory instances experiencing the enormous power of population.

> In the polygastric animalcules, spontaneous fission takes place so rapidly that it has been calculated by Prof. Ehrenberg that no fewer than 368 millions might be produced in a month from a single Paramecium; and even this astonishing rate of increase is far exceeded in another species, one individual of which [. . .], is calculated to generate 170 billions in four days.[62]

Nonetheless, the bulk of more recent discussions by unified growth economists seems to have systematically overlooked that the term "power of population" was merely intended to be used as a theoretical reference point that would only be realized under optimal environmental conditions, or – as an economist would call it today – under optimal economic incentives.

54 Mill (1848), book I, chapter X.
55 See, for example, Bloom et al. (2003).
56 See Livi-Bacci (2012), p. 12.
57 A variable is considered as stationary if its mean and variance do not change over time.
58 In a stationary population, changes in the birth/death rate accurately reflect changes in fertility/mortality and the inverse death rate displays life expectancy.
59 The natural population growth rate excludes net migration. Empirically, in the majority of states migration accounts for a small fraction of population growth only. It is therefore regarded as negligible in most works.
60 The reason for our inclination to meet these numbers with disbelief and skepticism might be rooted in our thinking being limited to changes which take place during our lifetime. After eighty years, the average individual would have generated an offspring of "merely" 1,000. However, as we are slow in observing gradual changes that last longer than a few generations, the effects of the subsequent 133 years are rarely taken into account and intuitively underestimated.
61 Herbert Spencer (1820–1903), British anthropologist, biologist, sociologist, subeditor for the journal The Economist, nominated for the Nobel Peace Prize in 1901 and the Nobel Prize in Literature in 1902 (declined).
62 Spencer (1852), §3.

3.3.2 The tendency to increase in numbers

Secondly, having thus stated the potential of exponential population growth, Malthus eventually suggested to what extent it was exerted in reality. Yet, by claiming that every population was observed to display the "tendency" to multiply in an exponential manner, he argued deductively, without examining the underlying motivation that induced each individual to generate progeny. It was left for C.R. Darwin[63], building on Malthus and Spencer, to remark that the origin of high individual rates of propagation was rooted in genetically varying inheritable traits:

> The fertility of each species will tend to increase, from the more fertile pairs producing a larger number of offspring, and these from their mere number will have the best chance of surviving, and will transmit their tendency to greater fertility.[64]

Hence, reproductive success might be seen as a dominant evolutionary strategy to every species, for if they did not conform to this rule they would, generation after generation, be reduced to a minor share of the population of the earth. As a result of this evolutionary process, every individual is with a high probability inherently equipped with a strong pursuit of procreation. In stating what economists would denominate a microeconomic theory of fertility behavior, Darwin argued that

> in looking at nature, it is most necessary [...] never to forget that every single organic being may be said to be striving to the utmost to increase in numbers.[65]

3.3.3 The limits to growth

A third premise of Malthus was offered by the quite incontrovertible statement that the space as well as the physical matter supplied by the Earth was limited. The limited space on Earth is unquestionably well–defined, and since extraterrestrial resources have not yet been accumulated in any considerable amount, we shall also agree on the finiteness of resources from the beginning of the existence of life until present times. Given that space is limited and presupposing that population consumes space, it is undeniable that there must exist some point at which population growth would have to come to a halt. More practically spoken, it should be obvious that there exists a limited amount of supply provided for the maintenance of all living beings that Malthus had, in the case of the human species, defined as "means of subsistence". From the existence of a limited resource constraint, in turn, he derived his first proposition that

> population is necessarily limited by the means of subsistence.[66]

---

[63] Charles Robert Darwin (1809–1882), British naturalist, geologist, biologist, founder of the theory of evolution by natural selection and sexual selection, fellow of the Royal Society.

[64] Darwin (1871), p. 319.

[65] Darwin (1859), chapter III.

[66] Malthus (1826), book I, chapter II.

3.3.4 Modeling the principle of population

Malthus combined the tendency for increase and the existence of a resource constraint to formulate the "principle of population" (in the following "PoP"):

> According to the principle of population, the human race has [...] a constant tendency to people a country fully up to the limits of subsistence; meaning, by these limits, the lowest quantity of food which will maintain a stationary population.[67]

To that effect, defining the means of subsistence as production Y and the average individual subsistence level as production per capita $y = Y/N$, the denominator of the latter would tend to rise until an economy was fully peopled up to the limits of subsistence, holding productivity roughly on a constant minimum subsistence level. Once an economy was fully peopled, changes in production are decisive in determining the growth rate of the population. In his second proposition, Malthus wrote unambiguously that

> population invariably increases, where the means of subsistence increase, [...].[68]

Combining this proposition with the power to increase, the PoP can also be formulated as the tendency of population to increase whenever production has been raised. Consequently, the relationship defining the PoP will subsequently be modeled through a positive proportional effect running from production to the number of births. While it is biologically evident that a positive wealth effect on the number of births cannot, on average, be realized earlier than nine months after a shock in production, and accounting for a lagged fertility decision of not more than one year, it is plausible to suspect the number of births to react on average at least one year after the shock in the means of subsistence took place, written for simplicity as $B_t = Y_{t-1}$. Again, as the natural population change consists of the difference between births and deaths, the principle of population may be modeled by using a law of population accumulation of the form

$$N_t = B_t - D_t + N_{t-1} = Y_{t-1} - d_t N_{t-1} + N_{t-1}.$$

Rewriting this equation in terms of per capita by dividing both sides by N_{t-1} gives

$$g_{N_t} = \frac{Y_{t-1}}{N_{t-1}} - d_t \frac{N_{t-1}}{N_{t-1}} \Leftrightarrow b_t - d_t = y_{t-1} - d_t \Leftrightarrow b_t = y_{t-1}.$$

Reformulating the last relation in terms of growth rates and logarithmizing both sides to receive a linear relationship yields

$$\frac{b_t}{b_{t-1}} = \frac{y_{t-1}}{y_{t-2}} \Leftrightarrow ln\ b_t \approx ln\ b_{t-1} \overbrace{+ g_{y_{t-1}}}^{PoP}. \tag{3.9}$$

We have thus preliminary modeled the principle of population as only component of the theory of population in such a way that changes in productivity have a positive impact on the birth rate of the form $\frac{\partial b_t}{\partial g_{y_{t-x}}} > 0$, with $x = 1$ accounting for a fertility lag.

[67] Malthus in Senior (1836), p. 147.
[68] Malthus (1826), book I, chapter I.

3.4 Classical Endogenous Growth Theory: The Malthusian Trap

> The Malthusian law of population is one of the great achievements of thought. Together with the principle of the division of labor it provided the foundations of modern biology and for the theory of evolution; [...] the objections raised against the Malthusian law as well as against the law of [diminishing] returns are vain and trivial. Both laws are indisputable.[69]

The best illustration and at the same time overwhelming evidence of the concerted operation of the three above defined principles is provided by their application to the theory of evolution by natural selection. Darwin as well as A.R. Wallace[70] developed their idea of natural selection independently from each other only after they had read Malthus' book on the principle of population.[71] Since the key driver of the evolutionary mechanism was in addition to genetic variation found in a constant tendency toward a state of overpopulation in which the least favored individuals were weeded out, the "survival of the fittest" was, according to Wallace, better characterized as the "extinction of the unfit".[72] Taking this process of evolution by overpopulation for granted, it must – at least at some early period – have been applicable to homo sapiens as well. Clark (2007) even concluded that at least until about 1800 AD, human and animal evolutions were governed by the same essential laws of population growth and that economic outcomes did not crucially differ, suggesting that human beings were not able to restrict their numbers consciously. Similarly, the principle of population was in classical economics held to be responsible for the Malthusian trap. North characterized this finding as the "big dark cloud above the enlightenment movement."[73] To better understand his conclusion, we will in this section provide illustrations to show that the Malthusian trap is established by the simultaneous operation of the PoDR, the PoLD and the PoP. To sustain the universality of the principles, we will start with the illustration of the non–human economy, exhibiting a low degree of labor division, before entering the subject of the human economy with a high degree of labor division.

3.4.1 The principles of population and of diminishing returns in the non-human economy

Firstly, to depict the concerted action of the principle of population together with the principle of diminishing returns, we will consider the most primitive case of a non–human economy with a fixed resource constraint, in which the inhabitants are assumed to be incapable of artificially increasing the means of subsistence. In addition, as it is often asserted that homo sapiens is the only species capable of birth control, the latter is equally supposed to be non–existent and will be examined at a later point after having considered the human economy. We will, therefore, turn to an economy where the principle of population is reflected by an unrestricted maximum

69 v. Mises (1949), p. 663.
70 Alfred Russel Wallace (1823–1913), British naturalist, co-founder of the theory of evolution by natural selection, fellow of the Royal Society.
71 "[...] I saw, on reading Malthus on Population, that natural selection was the inevitable result of the rapid increase of all organic beings [...]" Darwin (1868).
72 Wallace (1890), p. 337.
73 North (2013).

birth rate. In this simple case, the population pressure resulting from the principle of population could merely be released by a proportionally rising death rate.

Assuming the principle of population to have operated for millions of years, Darwin justly concluded that in reality, to secure survival, every established species must have occupied an economic niche providing subsistence.

> Owing to the high geometrical rate of increase of all organic beings, each area is already fully stocked with inhabitants.[74]

The presumed steady operation of an (unrestricted) exponential increase in numbers implied first and foremost that the emerging generation of a group of living beings tended to outnumber the former generation. However, since the supposedly stable environment did not provide additional niches for the upcoming generation, some individuals had to remain niche–less. As a result, there would of necessity arise competition between these abundant individuals, resulting in a "struggle for existence", which is one of the most consolidated findings in biology.[75] The following example illustrates a very simple and obvious case of the pressure arising from the principle of population.

In a forest that is fully covered by beeches, it is impossible for seeds to start growing until an existing tree has died off. If, on the other hand, an old tree has recently vanished and thus supplied a vacant spot under the sunlight, the free area will, according to the principle of population, soon be covered by seedlings. While growing up, however, each seedling will consume an increasing amount of space and resources until irreconcilable conflicts emerge, as it is physically impossible for all seedlings to grow up to a full tree. Although the precise outcome of these conflicts may be uncertain in general, they cannot be bypassed and reveal themselves through regular competition between individuals.

> Hence, as more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life. It is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms; for in this case there can be no artificial increase of food, and no prudential restraint from marriage. Although some species may be now increasing, more or less rapidly, in numbers, all cannot do so, for the world would not hold them.[76]

Even though we might allow animal and human populations to respond far more dynamically to these conflicts, they are nonetheless subjected to the same laws of competition derived from the principle of population and the principle of diminishing returns. According to Darwin, some of these inevitably redundant individuals were determined to die prematurely, ultimately by starvation, although among most species advanced mechanisms of high mortality prevailed such as disease, infanticide, suicide or homicide. Malthus defined these economies where an abundant part of the population was regularly swept away through high mortality as being *positively checked*. From his definition that

[74] Darwin (1859), chapter IV.
[75] See, for example, Weiner (1995).
[76] Darwin (1859), chapter III.

> the positive checks to population [...] include every cause [...] which in any degree contributes to shorten the natural duration of life,[77]

it becomes clear that the strength of the positive checks and the quantity of the death rate are measured by the same magnitude.[78] Whenever the positive checks operate powerfully, the death rate is high. Where the death rate is close to its minimum level and the average individual lives out its "natural duration of life", the positive checks are the weakest. However, when measuring the operation of the positive checks, it should be borne in mind that their existence alone does not prove the presence of a strong degree of population pressure. Where they are measured it is merely proven that population growth is kept below its maximum rate. Although a high death rate is a by–product and a good indicator, it is neither a sufficient nor a necessary condition for the existence of deadly competition resulting from the principle of population and the principle of diminishing returns. In summary, the economy here outlined will tend to exhibit a constantly high death rate that can be modeled as part of our theory of population by

$$d_t = d_{t-1} \tag{3.10}$$

What is more, it should already be noted that there are instances in which the struggle for existence does not necessarily follow from an excess of newly born individuals expanding beyond the subsistence provided for it. Competition might also be called into action after an already fully stocked territory has been struck by a diminution of natural conditions, lowering the resource base, or from an increasing population owing to improved conditions for survival such as the disappearance of predators or diseases lowering mortality (we will return to the corresponding form of competition in chapter five). Nonetheless, the principle of population remains the most regular driving force for competition, for if there were no tendency for the number of births to exceed the number of deaths, each territory would not categorically be fully stocked. Only from the steadily repeated application of this universal natural process of competition on a given amount of territory may we derive the rule

> that each [individual] lives by a struggle at some period of its life; that heavy destruction inevitably falls either on the young or old, during each generation or at recurrent intervals. Lighten any check, mitigate the destruction ever so little, and the number of the species will almost instantaneously increase to any amount.[79]

3.4.2 The Malthusian trap in the non–human economy

Secondly, to fully understand the process of the Malthusian trap that led to the cycle of misery and economic stagnation, we will add the idea of the principle of labor division to extend the above illustration of a non–human economy. Notwithstanding the assumption of a fixed resource boundary in the former section, we should not too hastily fall into the error of believing that

[77] Malthus (1826), book I, chapter II.
[78] As has been mentioned above, in a stationary population, the "natural duration of life" (life expectancy) corresponds to the inverted death rate.
[79] Darwin (1859), chapter III.

the limits of subsistence in fact remain constant in every animal economy. By analyzing the divergence into different animal species from a common ancestor and establishing a mechanism for evolutionary development, Darwin proposed a way by which the natural resource barrier would be steadily raised. Although deadly conflicts were the rule, the pressure of population comprised milder forms of competition, for example pushing individuals into niches that could not possibly be occupied by the former generation. Since the progeny of most species differed from its parental generation in genetic endowment, it happened to explore living spaces that were denied to its ancestors, as is illustrated in the following example.

One might imagine a rodent colony having initially fully populated the ground of a given territory. Arising from exponential population growth, an abundant number of young individuals might be pushed into an environment so far unsuitable for the common rodent. With this progeny displaying genetic variation, there may at some point appear a specimen endowed with the ability to climb trees, another to dive into water and a third to dig into the soil – abilities that were denied to the parent generation. If these specific abilities, by exploring new kinds of nutrition providing additional subsistence, were sufficient to sustain offspring, the specimen had created their own niches. Once they were established in these specialized niches, their growing number of offspring, displaying another large pool of variation, would again be subjected to competition. By the process of "natural selection," the abundant descendants unfit for survival were, generation by generation, frequently weeded out, while those displaying the highest genetic fitness under the prevailing conditions tended to propagate most rapidly. In this way, becoming ever more slightly adapted to the new environment, the species squirrel, otter and mole emerged.

Thus, Darwin had derived two important outcomes of a freely operating principle of population. Firstly, in the case of the animal economy, genetic variation and the operation of the two principles of population and of diminishing returns are critical in generating specialization and as a by–product to lift the natural resource constraint. Since the overall population of individuals N increased the number of available niches, the natural limits of subsistence Y must have been raised as well. Hence, by the simple means of population growth and variation, competition had not only generated new species, but had also created a "symbiosis of specialized species" by which the resource constraint was permanently elevated, which can account for parallel increases in population and production. Secondly, although genetic variation ameliorated the original individual's prospects for survival, it did not enhance the material situation of its respective descendants in the long run, since the speed of increase of the means of subsistence derived from specialization was clearly inferior to the speed of population growth. Individual specialization was merely intended to secure immediate survival, not to accumulate wealth, and the offspring of the first individual was in most cases not much better off than those living before the divergence of the species had started, holding productivity on the same subsistence level y over the course of evolutionary development. Thus, it is due to the supreme power of population in outperforming innovation by genetic variation that the mechanism of natural selection could endure a very long time without producing any permanent individual material gains. To Darwin, the struggle for existence, which is a logical implication of the second outcome, formed the fundament of the theory of evolution by natural selection. He unambiguously urged his disciples to realize that

> nothing is easier than to admit in words the truth of the universal struggle for life, or more difficult – at least I have found it so – than constantly to bear this conclusion in mind. Yet unless it be thoroughly engrained in the mind, the whole economy of nature, with every fact on distribution, rarity, abundance, extinction, and variation, will be dimly seen or quite misunderstood.[80]

We will in the following realize that the superior power of population can be equally assumed to neutralize the benefits from all those innovations commonly referred to as "technological progress", when introducing the operation of the principle of labor division into the human economy.

3.4.3 The principles of population, of diminishing returns and of labor division in the human economy

Although the operation of the three principles has been sufficiently proven by the existence of natural selection among non-human species and is widely accepted in natural sciences, its relevance for mankind is not rarely doubted. Assuming the validity of the above Darwinian process of innovation, the most regular critique Malthus' theory has faced over the last two hundred years was the argument that homo sapiens apparently possessed the ability to raise its natural resource constraint self-dependently without necessarily having to rely on slow and arbitrary genetic improvement. Malthus however, as well as any other classical economist, was fully aware of the fact that increasing production was a regular phenomenon accompanied by human population growth. They understood that growth of production was to the largest part owed to individual specialization and exchange based on Smith's "division of labor".

However, as will be shown subsequently, the emergence of the Smithian division of labor in human economies, or of what is today sometimes called "Smithian growth,"[81] is not much different from what we have observed in the animal economy in the form of a symbiosis of specialized species. As was the case in the animal economy, the process of human specialization into different professions might be traced back to the competition derived from the operation of the principles of population and diminishing returns. As we proceed to presume that the preventive checks are non–existent and that fertility is exerted at its maximum level, we continue to argue that a newly emerging generation will tend to outnumber their foregoing cohorts, creating population pressure, competition and thus conflicts due to diminishing returns to population. In analogy to the case of the animal economy, this pressure of population would induce abundant individuals to explore new methods of production as is exemplified in the following.

Starting out as hunter and gatherer communities, those abundant members of a tribe deemed redundant by the community tended to venture capturing new species of prey or testing unknown fruits. If the exploration was unsuccessful, the respective individual would ultimately be exterminated. If it was successful, the new way of production could be permanently integrated into the overall production of the community, securing an additional niche for survival and again providing subsistence for further progeny. As with the tendency for growth, the number of suc-

[80] Darwin (1859), chapter III.
[81] See, for example, Kelly (1997).

cessful explorations steadily increased by trial and error, the community tended to accumulate numerous forms of production.

Notwithstanding these similarities to the animal economy, the mechanism by which specialized professions were accumulated seems to have been largely independent of genetic variation in the human economy. That the new processes were indeed regularly integrated into the economic system was, according to Smith, owed to the inherent and apparently unique tendency of human beings to "exchange" their products. The introduction of exchange and the correspondingly increasing demand brought with it the obvious advantage of "economies of scale" – to specialize in the production of one good and to supply the demand for the whole community, thus creating a division of labor among the working population. As long as such employment was sufficient to provide subsistence for a family, it could be properly denominated "profession."[82] It does not require a large degree of abstraction to imagine this evolutionary process to be responsible for every subsequently emerging profession, gradually diffusing from the gatherer to the rice farmer to the watchmaker up to the modern era. Smith used the production of the woolen coat as an example to demonstrate to what extent the division of labor and specialization had created a "symbiosis" in the human economy during preindustrial times.

> The shepherd, the sorter of the wool, the wool–comber or carder, the dyer, the scribbler, the spinner, the weaver, the fuller, the dresser, with many others [...] how many ship–builders, sailors, sail–makers, rope–makers, must have been employed to bring together the different drugs made use of [...] let us consider only what a variety of labour is requisite in order to form [...] the shears with which the shepherd clips the wool. The miner, the builder of the furnace for smelting the ore, the seller of the timber, the burner of the charcoal to be made use of in the smelting-house, the brick-maker, the brick-layer, the workmen who attend the furnace, the mill-wright, the forger, the smith, must all of them join their different arts in order to produce them.[83]

As with the animal economy, the same two important rules could be derived if the here illustrated human economy was empirically confirmed. The first rule being the idea that the combination of the principle of population, the principle of diminishing returns and specialization might have constituted the only source of permanent economic innovation and the second being the tendency to return to a subsistence level of productivity, since the speed of generating new innovations seems to have lagged behind the speed of population growth in most preindustrial economies, preventing production per capita from sustainably increasing. The latter point would certainly not come as a surprise as the growth of population is regarded to be the primary stimulus to innovations, for if population would not have kept up with production, there would have been no strong degree of competition. Indeed, following Mill's assessment, most classical economists were

[82] This is not to say that the Darwinian process of specialization had vanished. While still facing competition arising from the principle of population and thus constantly being forced to improve their productivity, the members of the community were in the long run determined to focus again on those processes that corresponded most efficiently to their individual natural endowments - a tendency that might be denoted as "Ricardian growth". This tendency to redistribute labor according to genetic ability is perhaps best illustrated by the sexual division of labor prevailing in many aboriginal societies where hunting is largely conducted by the males and gathering by the females. "Smithian growth", in contrast, is independent of the individual natural endowment.

[83] Smith (1776), book I, chapter I.

convinced that the tendency toward innovations improving economic output generally exhibited in a human economy must be fully owed to this kind of competition derived from the cooperation of the principles of labor divsion, of population and of diminishing returns.[84] The tendency for competition and to correspondingly specialize was to Spencer – as it was to Smith, Malthus and Darwin – the chief source of what he very generally termed "progress":

> From the beginning, pressure of population has been the proximate cause of progress. It produced the original diffusion of the race. It compelled men to abandon predatory habits and take to agriculture. It led to the clearing of the earth's surface. It forced men into the social state; made social organization inevitable; and has developed the social sentiments. It has stimulated to progressive improvements in production, and to increased skill and intelligence. It is daily pressing us into closer contact and more mutually dependent relationships.[85]

In summary, the three illustrated classical principles are assumed to have been working toward growth in population and production as well as constant productivity during the regime of stagnation. Hence, adhering to Galor's Malthusian trap, we observe a cycle of misery between population and production of the form

$$g_y = \overbrace{g_Y(g_N+)}^{PoLD} \underbrace{- \overbrace{g_N(g_Y+)}^{PoP}}_{PoDR} \approx 0 \tag{3.11}$$

3.4.4 Evidence of the Malthusian trap

As we have seen, the most striking logical argument in favor of the here outlined simultaneous operation of the three principles is provided by the establishment of the theory of evolution by way of natural selection. If more individuals had not been produced than could possibly survive, nature could not have weeded out the "unfit", since every individual would have simply survived and passed on their niche to the next generation. In this case, the genetic pool would have stayed unchanged and specialization may not ever have occured.

In addition, from a historical point of view, there are three facts that hint at once strongly at the functioning of the above evolutionary model of economic growth in the human economy up until the year 1800. Firstly, most economic historians will concur with the stylized fact that human production per capita did not crucially differ in the year 1800 as compared to the year 10,000 BC.[86] Secondly, it has been estimated that, although with no inconsiderable oscillations, the human population rose exponentially from roughly six million to about 1,000 million over the same time span.[87] Thirdly, presuming in addition that, as with every species, the Earth had already been "fully stocked" with human individuals in the first place, it is obvious that an increase in professions took place over the same period.[88] From the last point it seems proven

84 "Only through the principle of competition has political economy any pretension to the character of a science." Mill (1848), book II, chapter IV.
85 Spencer (1852), §16.
86 See, for example, McCloskey (2010), p. 2.
87 See Livi-Bacci (2012), p. 25.
88 If the number of professions had not changed, we would today witness a population of eight billion hunters and gatherers.

that specialization had occurred. Since, however, productivity had not increased in the long run although specialization had lifted the resource constraint, it is evident that population growth must have fully consumed the gains from specialization. This deduction represents the logic of the "Malthusian trap" as it was intended by Malthus in his original "essay" to form a cycle of misery.

As a final point of evidence, Malthus' original notion that the number of niches created by specialization is regarded to be insufficient to provide the emerging generations with employment is well supported by high rates of mortality and emigration that often accompanied high growth rates in population. More specifically, when turning to the testified pattern of historical populations, innumerable instances can be found in which the negative effects of the principle of diminishing returns outweighed the positive effects of the principle of labor division due to an unrestricted principle of population, inducing concomitant mortality crises and migration. This is, for example, well documented for the Viking and Mongolian expansions[89], the Crusades[90] or the Irish Famine 1846–1847[91] and plausibly depicted for the case of native American civilizations[92] Likewise, McCulloch observed this form of "overpopulation" as a regular historical phenomenon, stating that

> wars, plagues, and famines, those "terrible correctives" [. . .] of the redundance of mankind, set the operation of [diminishing returns] in a striking point of view. They lessen the number of the inhabitants, without, in most cases, proportionally lessening the capital that feeds and maintains them.[93]

3.4.5 Simulation of the Malthusian trap

Having stated the established and classical literature and having found sufficient evidence for the existence of the three classical principles forming the Malthusian trap, the structural equations of the neoclassical theory of production and the Malthusian theory of population are mathematically summarized in this subsection to form a simulation that can illustrate the operation of the Malthusian trap in accordance with the concomitant stylized facts of the cycle of misery and

---

89 "[F]or we are aware that the hordes of Central Asia and of the Northern parts of Europe, and the surplus inhabitants of some small communities, such as the petty States of ancient Greece and Phoenicia, appear to have found, the one in colonization, the others in armed migrations, a periodical outlet;" Senior (1836), p. 41.

90 "Your land is shut in on all sides by the sea and mountains, and is too thickly populated. There is not much wealth here, and the soil scarcely yields enough to support you. On this account you kill and devour each other, and carry on war and mutually destroy each other. Let your hatred and quarrels cease, your civil wars come to an end, and all your dissensions stop. Set out on the road to the holy sepulcher, take the land from that wicked people, and make it your own. That land which, as the Scripture says, is flowing with milk and honey, God gave to the children of Israel." Pope Urban II. speech at the council of Clermont (1095).

91 "[T]he recent condition of Ireland and of the Highlands (...) ought not be regarded merely as a visitation of Providence, calling for temporary aid from the rest of the nation, but as an indication of a previously unsound condition of the population, [...]". Alison (1847), p. 9.

92 "In fact, one of the main lesson to be learned from the collapses of the Maya, Anasazi, Easter Islanders, and those other past societies [...] is simple: maximum population, wealth, resource consumption, and waste production mean maximum environmental impact, approaching the limit where impact outstrips resources." Diamond (2005), p. 509.

93 McCulloch (1863), part I, chapter VIII.

economic stagnation. The resulting system of linear equations provided by the above derived equations 3.7, 3.9, 3.10 can be written as

$$\begin{aligned} g_{y_t} &= \alpha_1 ln\ b_{t-15} - \alpha_2 ln\ b_t \\ ln\ b_t &= \alpha_3 ln\ b_{t-1} + \alpha_4 g_{y_{t-1}} \ , \\ d_t &= \alpha_5 d_{t-1} \end{aligned} \tag{3.12}$$

where two additional assumptions have been made to arrive at this system. Firstly, the length of one generation is reduced to fifteen years, which seems also to be the lowest plausible average age at which a regular employment is secured in the human economy. Secondly, leaving some room for the interpretation of the relative operation of the principles over time, the magnitude of each effect is represented by an undefined (potentially time–varying) coefficient. A full econometric account for the significance of the three principles based on this system of equations will be given in chapter seven. At present, however, we are satisfied with a simulation of a very simple model, where all parameters are set to unity. Moreover, for illustrative purposes we leave aside the exponential character of the relation between birth rate and productivity growth. This merely implies that the effect of birth rate on productivity and the positive effect of productivity on birth rate are weaker in the simulation than they ought to be found in reality.

$$\begin{aligned} g_{yt} &= \overbrace{\alpha_1 b_{t-15}}^{PoLD} \overbrace{-\alpha_2 b_t}^{PoDR} + \epsilon_{g_{y15}} \\ b_t &= \alpha_3 b_{t-1} + \overbrace{\alpha_4 g_{y_{t-1}}}^{PoP} \\ d_t &= \alpha_5 d_{t-1} \end{aligned} \tag{3.13}$$

Table 3.1: Calibration of the system of 3.13

α_1	α_2	α_3	α_4	α_5	b_0	d_0	g_{y0}	$\epsilon_{g_{y15}}$
1	1	1	1	1	0.035	0.020	0.000	0.025

Calibrating the parameters and initial values according to 3.1 and simulating a shock in productivity after 15 periods yields Figure 3.3, the Malthusian trap. More explicitly, shocking the growth rate of productivity in period fifteen raises the birth rate one period later owing to the PoP. This increase in population instantly consumes the former gains in productivity due to the PoDR. Hereafter, fifteen periods of stagnation follow until the larger birth cohort has come of age to participate in the labor market, thereby increasing productivity growth via the PoLD, resulting in a further increase in births and so forth. Over time, as is illustrated in Figure 3.4, this short–run mechanism leads to a steady increase in the level of production and population, whereas the growth rates as well as productivity are observed to be relatively stable over the long run. Consequently, our model can also account for the recorded stylized facts of the cycle of misery and economic stagnation. An overview of the theoretical findings of Part I of this work is provided by Figure 3.5.

Figure 3.3: A simulation of the mechanism of stagnation in growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

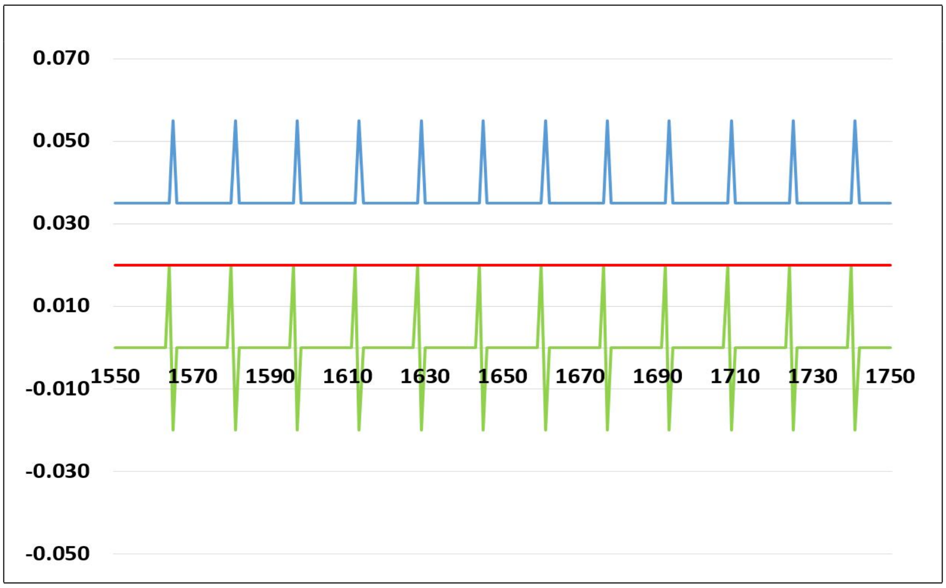


Figure 3.4: A simulation of the mechanism of stagnation in level variables (population (orange), production (blue) and GDP per capita (green)).

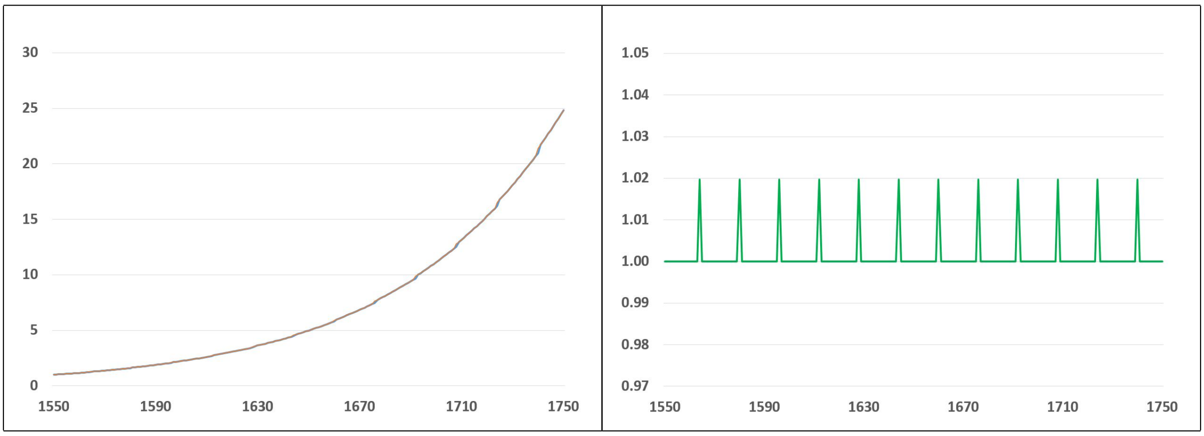
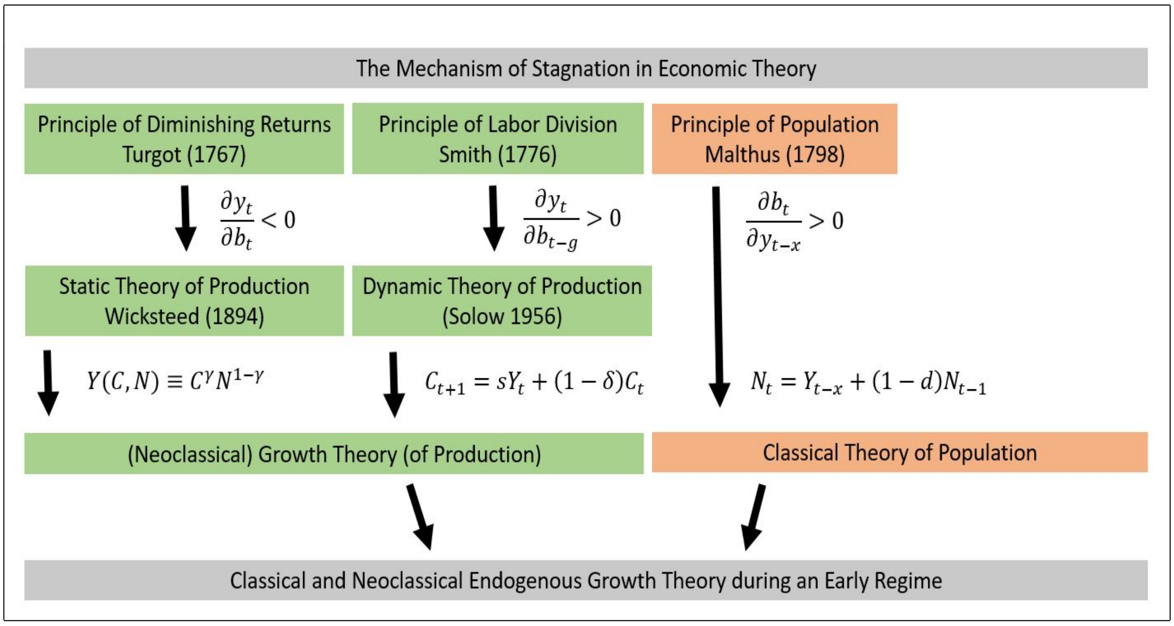


Figure 3.5: Theoretical findings of Part I.



Part II

The Era of Economic Development

Chapter 4

The Regime of Economic Development

> This law [of population] cannot be denied by anybody. I could list as many referees supporting it as there are great and prominent names in the science of economics, and in particular, of the [classical] school itself, since it is precisely this [classical] school, that has uncovered and established this law. This iron and cruel law, gentlemen, you have to deeply, deeply engrain into your mind and use as a reference point on your entire thinking.
>
> At this occasion, I will provide you and the whole laboring profession with an infallible way of arguing to escape all of those misdirections and delusions once and for all. You will have to confront every man who is talking about the improvement of the situation of the laboring profession with the following question: Whether he accepts the law or not. If he does not so, you will have to presume that this man either tries to deceive you or that he is of the most miserable inexperience in the science of economics. Since there exists, as I have already mentioned, in the [classical] school itself not even one prominent economist denying this law. Adam Smith as well as Say, Ricardo as well as Malthus, Bastiat as well as John Stuart Mill uniformly agree in acknowledging it. Here, agreement prevails between all men of the science.
>
> Now, if this man, speaking about the situation of the laboring profession, has answered to your question to accept the law, you ask further: How does he intend to release it? And if he does not know an answer, you may calmly turn your back on him. He is an empty talker, who wants to deceive and blind you or himself with empty phrases.[1]

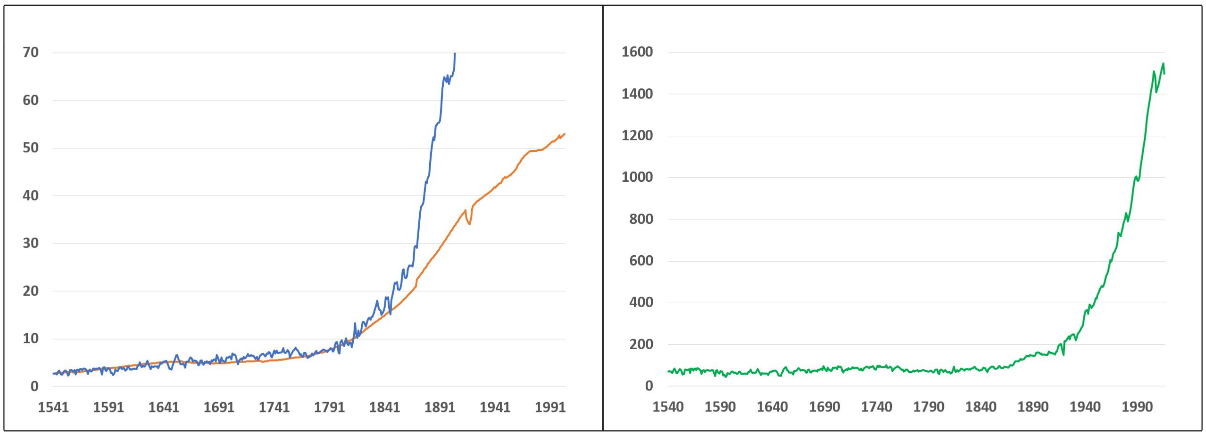
4.1 Evolution of GDP per Capita and GDP

In this chapter, it will be attempted to reveal the effects responsible for releasing the pressure of population and for the successively observed increase in GDP per capita – the breakout from stagnation. The analysis of the regime of development starts with a graphical examination of the British data series employed in chapter two, extended to the year 2016 by using official data sources provided by Mitchell (2013) and the Worldbank (2018). In Appendix 11.2, examples are depicted to confirm the impression that the British stylized facts also reliably account for a global generalization. The right graph of Figure 4.1 shows the evolution of British GDP per capita. A first glance reveals that GDP per capita slowly rose during the nineteenth century

[1] Lassalle (1863).

and began to rapidly increase during the twentieth century. Having introduced the regime of stagnation and notwithstanding the correctness of the considerations made in the last chapter, it is evident that the interpretation of a "Malthusian trap" cannot be upheld after the year 1800, as the cycle of misery broke down (left graph of Figure 4.1). Today, the pressure of population as formulated by the classical economists plays almost no role in economic policies of developed countries any longer.

Figure 4.1: Divergence of GDP (left, blue) and population size (left, orange) in Britain 1541–2014. GDP per capita (right, green) in Britain 1541–2014.



Sources: GDP: Clark (2009) for 1541–1871, Mitchell (2013) for 1871–2010, Population: Wrigley and Schofield (1981) for 1541–1871, Mitchell (2013) for 1871–2010.

To analyze the breakout from stagnation in more detail, the time series of GDP and population displayed in the left graph of Figure 4.1 provides important additional information. Firstly, the year 1847, which was characterized by economic depression and widespread famines throughout Europe – eventually causing equally widespread political unrest during the year 1848 – seems to be the latest possible date at which we may speak of stagnation in GDP per capita, since the ratio between GDP and population is found to be below the level of the year 1541 for the last time. Accordingly, the year 1848 should be chosen as the latest possible date for the initiation of the British breakout. Secondly, GDP and population appear to have accelerated during the period 1760–1815. Even though the graph does not capture the complete series, the acceleration of GDP continued well into the 20th century and has slowed down only recently. Thirdly, population growth started to slow down during the beginning of the nineteenth century.

Subsequently, the underlying forces that caused GDP per capita to increase will be investigated by analyzing the evolution of GDP and population separately. The author has identified three major fallacies of currently circulating economic theories that deserve a more explicit clarification, as they continue to disguise the real cause for and to prevent a proper understanding of the breakout from stagnation. To enlighten these fallacies and to advance a resuscitation of the classical framework, the next sections will explore and criticize currently and formerly established theories of growth and development, before returning to the classical position.

4.2 Fallacy 1: The Conventional Wisdom that an Increase in Production Produced the Breakout from Stagnation

4.2.1 An endogenous shock in production?

We have seen that Kremer's idea of population being generally capable of raising production more than proportionally cannot be supported by the British data until around 1800 AD. We will now briefly examine that even after the year 1800, the notion of "increasing returns to population by way of technological progress" does not stand the test of reality and can therefore not be regarded as a general economic principle during the era of development either.

Naturally, the simplest way to explain a lasting increase in production per capita is to claim that production possessed the power to outgrow population. Although British average production gains generated during the period 1760–1820 had been regularly outperformed by population growth, Smith's considerations had already influenced some economists to believe that the benefits stemming from the division of labor had the potential to outperform increases in population. This view had already attracted widespread sympathies during the early nineteenth century:

> On the one side are those who believe that an increase of numbers is necessarily accompanied not merely by a positive, but by a relative increase of productive power; that density of population is the cause and the test of prosperity; and that, were every nation under the sun to be released from all the natural and artificial checks on their increase, and to start of breeding at the fastest possible rate, many, very many generations must elapse before any necessary pressure could be felt.[2]

Confronted with this argument and being very well acquainted with the process of labor division, specialization and therefore "technological progress", Malthus defended the idea that the power of population *growth* was superior to the power of *growth* in production:

> The power of the earth to produce subsistence is certainly not unlimited, but it is strictly speaking indefinite; that is, its limits are not defined, and the time will probably never arrive when we shall be able to say, that no further labour or ingenuity of man could make further additions to it. But the power of obtaining an additional quantity of [subsistence] from the earth by proper management, and in a certain time, has the most remote relation imaginable to the power of keeping pace with an unrestricted increase of population.[3]

Employing a simple illustration, H. C. Carey[4] (1837) equally hinted at the fact that the principle of diminishing returns to population would ultimately prevail:

> If land would always yield in proportion to the quantity of [population] applied to it, there would be no need to cultivate more than a single farm, or a single district, for the supply of any number of inhabitants; and because such cannot be the case, it is assumed that every fresh application of [population] to cultivation, must be attended with a diminished return.[5]

[2] Senior (1836), p. 146.
[3] Malthus (1826), book V, chapter I.
[4] Henry Charles Carey (1793–1879), US–American economist, chief economic adviser to US president Abraham Lincoln.
[5] Carey (1837), vol. 3, p. 8.

As a matter of fact, it ought to appear utterly impossible to the modern economist to supply a population of the aforementioned potential of around 135,000 trillion inhabitants of the earth that arose from an unrestricted growth in population within about two hundred years even under the most favourable conditions for specialization and technological progress. In spite of those exceptional historical instances in which the discovery of new territory or of rare natural resources raised the production of an economy tremendously over the short run, Senior (1836) argued likewise that this cannot be the permanent state of affairs.

> Although, therefore, it is not possible to assign any certain limits to the progress of improvement, it is notwithstanding evident that it cannot continue for any considerable period to advance in the same proportion that population would advance supposing [resources] were abundantly supplied.[6]

Furthermore, while it is observed that the annual long–run growth rate of production rarely exceeded five or six percent in historically recorded economies, we have shown that population possessed the ability to grow by around nine percent annually. Accordingly, we would have to expect a permanent growth rate of more than nine percent in those economies that have still succeeded in departing from a subsistence level to justify the idea that production had outrun population. Consequently, as it is neither theoretically nor empirically convincing, the doctrine that population growth would generally raise production more than proportionally through faster accumulation of "technology" cannot constitute an economic principle.

4.2.2 An exogenous shock in production?

Often, the take–off date for the increase in productivity is observed to closely correspond to the beginning of the so–called "industrial revolution"[7]. This coincidence has often been viewed, in particular by historians, as a turning point in the history of mankind toward a new path of sustained economic development. What is more, it led to the conventional wisdom among economic historians that the industrial revolution, in the form of rapid "technological progress", even *caused* the breakout from stagnation.[8] To be sure, since it is almost identical with a lasting increase in the human resource constraint, "technology" was certainly fundamental in inducing the British industrial revolution. However, if we continued to argue that "technology" is simply caused by population growth, we would have to point again to an endogenous shock in production, which has already been falsified in the former subsection.

To circumvent the endogeneity problem, contemporaneous economic historians devoted their efforts to a clearer understanding of the original emergence of an "exogenous shock" as causing the British industrial revolution. For example, Clark (2007) proposed a unique British social upward mobility, by which richer individuals had more offspring and spread their hard–working attitudes by passing it on to their children. Allen (2009) suggested that increasing British *nominal* wages allowed for rapid capital substitution and technological progress, whereas Mokyr

6 Senior (1836), p. 147.
7 A term that might need some reassessment, as it is, for scientific purposes, not sufficiently well defined.
8 See, for example, Allen (2009).

(2016) ascribed the emergence of this "technological revolution" to a unique British culture of enlightenment. Each of these causes seem to be assumed to have raised production tremendously by inducing an industrial revolution, causing the breakout from stagnation.

To avoid a common misunderstanding, on this occasion the point must be stressed that the exaggerated role the industrial revolution played in explaining the transition to economic development is arguably rooted in the ambiguous use of the term "economic growth". However, the industrial revolution as well as the term "technology" both refer to growth in terms of *total* GDP Y only. They can much less account for the increase in GDP *per capita* y, since its observed evolution in the eigteenth century is well in line with the hitherto regular pattern of a Malthusian trap, as is illustrated in the next paragraph.

Beginning in the eighteenth century, aided chiefly by the introduction of the potato and the disappearance of the plague epidemic, European economies started to experience strong population growth.[9] Since an increasingly growing population meant an increasingly larger number of innovations from trial and error, the pace of specialization increased – a process that would culminate in the industrial revolution, which was again an important reference point for Smith's considerations on the division of labor. However, the emergence of a process of industrialization, as is often drawn by historians in connection with the process of urbanization, did *not* overcome the regime of stagnation but seems to have followed the traditional cycle of misery. The industrial revolution was merely another exploration of new production tasks resulting from the competition of a growing labor force, as in the case of the hunter–gatherer society. When the countryside became ever more densely (over–)populated, the markets and harbors of first villages, then towns and ultimately cities represented a natural resort for absorbing the abundant farm workers into specialized factories and transport companies making use of economies of scale without, however, raising productivity of the average laborer in its early stages.[10] On the contrary, the industrial revolution most arguably even aggravated the population pressure by providing subsistence for an additional population, further stimulating the population explosion that culminated in the year 1815 in a population growth rate of 1.8%. Determining the beginning of the first British industrial revolution to the late 1760s, as is conventionally done, we observe parallel growth in GDP (75%) and in population (78%) during the first fifty years (1770–1820), meaning that GDP per capita stagnated or even declined during this period.

As a consequence, the only substantial difference between the process of the first British industrial revolution and the process of the neolithic revolution seems to have consisted in the speed they exhibited in spreading innovations due to a varying total population increase, while both events were subject to the same underlying Malthusian and Smithian principles. Another indication that the Malthusian trap was still well in existence during the industrial revolution is given by an international comparison: Even today we observe in most industrializing, emerging economies that concomitant strong population pressure tends to force abundant individuals to emigrate or to drive them into deadly competition, often by collectively waging war.[11]

[9] See Nunn and Qian (2011) on the potato, Langer (1963) on the plague.

[10] According to Allen (2001) and Clark (2009), a lasting increase in English wages cannot be observed until after 1820.

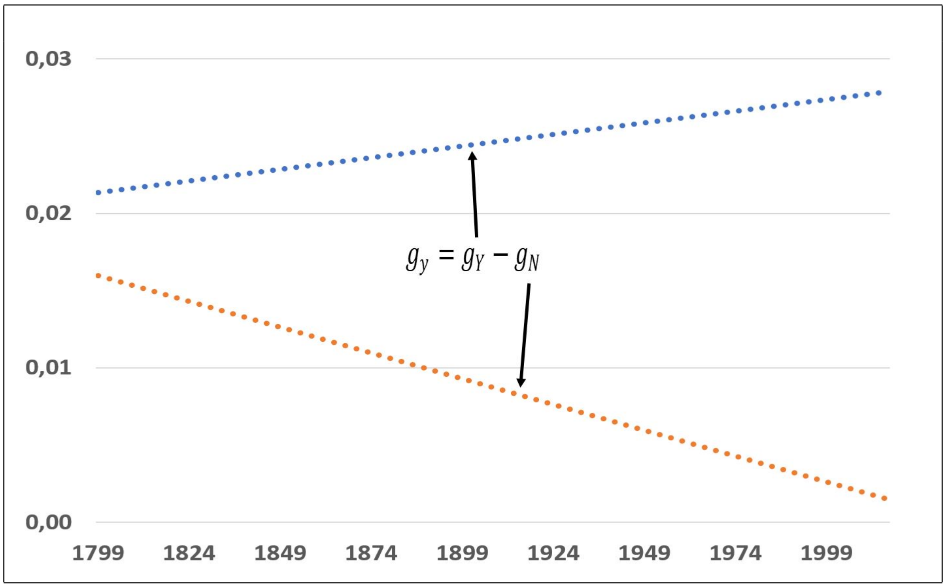
[11] Korotayev (2011).

In addition to the above considered theories employing an exogenous shock in production, there exist countless further attempts to justify the breakout from stagnation by stating that "countries have produced more" of which we may merely list the currently two most popular ones. While Acemoglu and Robinson (2012) follow the institutional approach of North (1992), by which proper political and economic institutions might have triggered the breakout from stagnation, McCloskey (2010), criticizing all of the above approaches, suggested that "economics can't explain the modern world" before proposing that British "burgeois dignity" may have induced the breakout. Apart from the fact that these approaches are rather ad hoc theories and therefore do not qualify for an endogenous growth theory, they all lack one crucial insight: The breakout from stagnation was not decisively triggered by "better ways of production", which include better ways of political organization. *The increase in GDP per capita was chiefly caused by a deceleration in the denominator population.* Although most of the above authors accept the existence of a Malthusian population trap at some earlier point in time, almost all of these theories do not attempt to explain why population would, from the nineteenth century onwards, fail to catch up with production. However, as we will see in the following section, a framework of a growth regime cannot simply disregard the mechanism of stagnation, but must incorporate its principles and explain why population growth slowed down, if it was intended to yield a unified growth theory.

4.3 Evolution of GDP per Capita, GDP, Population and Birth Rate

If we date the beginning of the British industrial revolution to the introduction of Watt's steam engine in 1769, it clearly preceded the British demographic slowdown, the first sign of which cannot be detected before the year 1815.[12] Moreover, following Clark and Allen, we may identify the period during which the take-off must have taken place to lie somewhere in the period 1815–1848. Challenging the conventional wisdom, since the demographic slowdown coincides more closely with the breakout from stagnation than the industrial revolution, the former seems to be the more suitable candidate as a cause for growth. This is confirmed when displaying growth rates of GDP and population instead of level variables over the last two centuries, as has been done in Figure 4.2. Here, the trended GDP per capita growth can easily be derived as the approximate difference between the linear trends of GDP growth and population growth. As we observe a divergence between population (orange) and production growth (blue), GDP per capita growth becomes increasingly positive. Moreover, one finds the deceleration of population growth to have about *two to three times as much explanatory power* compared to the acceleration of GDP growth. Accordingly, the largest share of the increase in GDP per capita growth must be attributed to the demographic slowdown and much less to an (howsoever defined) industrial revolution.

Figure 4.2: Divergence of British GDP growth (blue) and population growth (orange) 1800–2010.

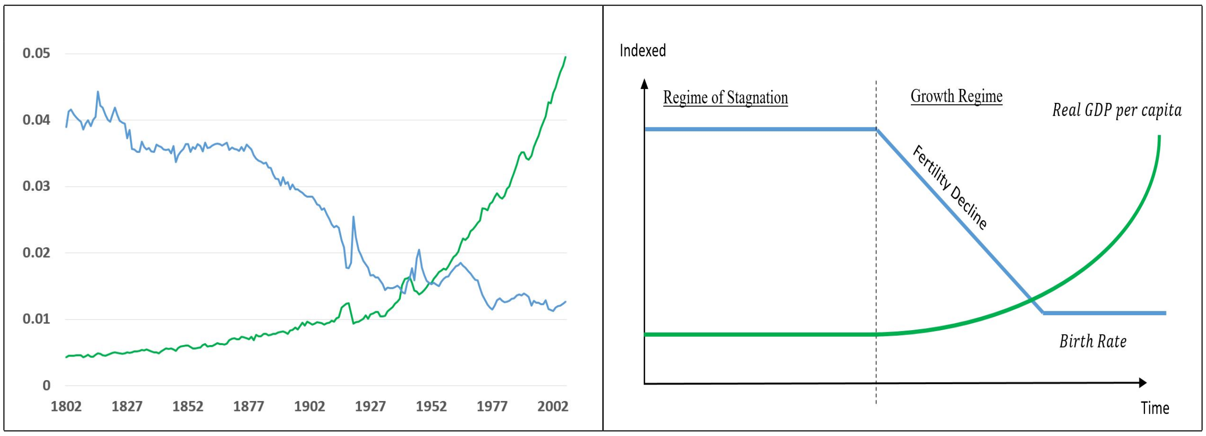


Sources: GDP: Clark (2009) for 1800–1871, Mitchell (2013) for 1871–2010, Population: Wrigley and Schofield (1981) for 1800–1871, Mitchell (2013) for 1871–2010.

[12] Alter and Clark conclude that "the Industrial Revolution and the Demographic Transition are the two great forces that explain the upward march of modern incomes. So far they have stood independently, the Industrial Revolution preceding the [Demographic Transition] in Europe by more than 100 years." Alter and Clark (2010), p. 44.

Since we have already seen that the variable (natural) population growth can be exhaustively split up into birth rates and death rates and since, as we will see further below, both vital rates display relatively continuous negative linear trends after the year 1815, the decline in population growth must be solely assigned to the falling birth rate – a process that has been named "fertility decline". As part of the population slowdown, the latter must by definition have, to a certain degree, been associated with increasing GDP per capita. As is illustrated in Figure 4.3, the British fertility decline accompanied the increase in GDP per capita after around 1800. When employing Mitchell's (2013) international historical statistics, this idea is confirmed by recognizing that no signs can be found of any economy in which a sustained rise in GDP per capita can be detected before a sustained decline in birth rates has been realized (see again Appendix 11.2 for a number of examples). Thus, the stylized fact "cross of wealth" pictured in the right graph of Figure 4.3 is assumed to properly display the general pattern between birth rate and GDP per capita.

Figure 4.3: Left graph: British "cross of wealth": Birth rate (blue) and GDP per capita (green) 1802–2007. Right graph: Stylized fact "cross of wealth".



Sources: GDP: Clark (2009) for 1800–1871, Mitchell (2013) for 1871–2010, Birth Rate: Wrigley and Schofield (1981) for 1800–1871, Mitchell (2013) for 1871–2010.

4.4 The Regime of Economic Development: Stylized Facts and Theory

The most comprehensive recent elaboration on the stylized facts of growth is probably summarized in Galor (2011), who suspects a causal link between the deceleration of the birth rate and the breakout from stagnation. In summary, Galor's unified growth theory builds on the following observed stylized facts of economic development:

1. Every economy was at some point over the past three centuries caught in a *regime of stagnation*, where productivity remained at a low level and birth rates at a high level.

2. Today, almost all economies have left the regime of stagnation in favor of a *growth regime*, where productivity increases from a low level to a high level.

3. Roughly at the same time as these economies left the regime of stagnation, a *fertility decline* set in, by which the birth rate declined from a high level to a low level.[13]

Having established a negative relationship between birth rate and GDP per capita growth, it is the task of the growth economist to model hypotheses incorporating causal effects. Advocates of the Kremerian model tend to neglect the fertility decline, maintaining the conventional view that production "increased sufficiently rapidly not only to outpace population growth, but to overcome diminishing returns,"[14] which is no longer tenable, given the considerations made in the former sections. Proponents of Galor's unified growth theory must object to this assessment, as the data show that more mitigated population growth was responsible for the divergence between GDP and population. However, if a fertility decline is globally found to coincide with a GDP per capita take–off, the question is frequently raised whether the transition to growth was an immediate consequence of or an immediate cause for the fertility decline. While the second hypothesis will be discussed in section 4.7, demographers tend to treat the fertility decline rather as a cause for an increase in productivity, inverting the concept of stagnation as follows. If initially strong population growth has been shown to impede increases in living standards, a deceleration of population growth opens up the possibility of modern economic development.[15] In the next two sections, it will be argued that this demographic predicition is well in line with classical theory.

4.5 Fallacy 2: The conventional wisdom that the principle of population has been falsified

Two hundred years ago, Malthus' principle of population was widely–known and its importance with regard to economic theory seemed, although controversial among laymen, generally accepted among economists. During the nineteenth century, it constituted the theoretical foundation not only of the science of political economy, but also of the emerging sciences of sociology and biology. However, over the course of the centuries, since new generations were not confronted with the same everyday problems the classical economists were facing, its popularity declined sharply as it was first increasingly misinterpreted and finally considered to have been falsified. The controversial and famous argument Malthus had brought up was to conjecture that the principle of population would in *reality inevitably* induce population, rising exponentially ("geometrically"), to permanently catch up to any higher level of production, which rose merely linearly ("arithmetically"), for Malthus (1798) had written in his original essay that

> natural inequality of the two powers of population and of production in the earth, and that great law of our nature which must constantly keep their effects equal, form the great difficulty that to me appears insurmountable in the way to the perfectibility of society. All other arguments are of slight and subordinate consideration in comparison of this. I see no way by which man can escape from the weight of this law which pervades all animated nature.[16]

[13] See Thompson (1930), who observed the fertility transition as part of the "demographic transition."
[14] See Komlos (2003).
[15] See Livi-Bacci (2012).
[16] Malthus (1798), chapter I.

Against this statement, an important criticism regarding the "great difficulty that appears unsurmountable" was legitimately raised. Even today, in spite of the illustrated early stagnation of GDP per capita and the recorded evidence in many animal species, the notion of the inevitability of a regime of stagnation does not attract too many economists for the following two reasons: Firstly, they observe globally increasing GDP per capita. Secondly, they observe at the same time decreasing population growth.

However, in his later editions Malthus became more optimistic and clearly admitted that his original version of an "unrestricted" principle of population inevitably resulting in a permanent regime of stagnation was misleading. Having travelled within large parts of Europe, gathering impressions and population data, he arrived at the insight that it was possible to embank the exponential power of population, attenuating his former conclusions in his later editions (1803-1826) by employing more frequently the term "tendency". Accordingly, the principle of population was to be interpreted as a permanently operating, abstract tendency (using the word tendency to express a propensity toward an increase in numbers) employed as a reference point for theoretical considerations. The actual historical era of stagnation, in contrast, had to be viewed as a readily testable empirical fact, employed as a practical benchmark on real observations. Hence, the principle of population had to be accepted as a fixed "law", like the "law of gravity", whereas the mechanism of a Malthusian trap and the resulting regime of stagnation could be circumvented under proper conditions. Consequently, as it would sometimes not reveal itself at first glance, the operation of the principle of population alone does *not* require every population to exhibit exponential growth in reality at all times, as is sometimes asserted, but rather reflects a latent "pressure" steadily operating toward an increase of numbers. Among others, McCulloch (1863) sustained the universality of the principle, maintaining that humanity had indeed been facing the principle of population at any point in history:

> The principle, whose operation under favourable circumstances has thus developed itself, is, in the language of geometers, a "constant" quantity. The same power that has doubled the population of Kentucky, Illinois, and New South Wales in five-and-twenty or thirty years, exists everywhere, and is equally energetic in England, France, and Holland.[17]

Notwithstanding Malthus' reconsideration, the majority of modern economists seems to stay intellectually trapped in his first essay on population, inclined to put the "tendency" of the principle of population on the same level with a "self–evident fact", the era of stagnation.[18] Senior had already perceived a widespread ignorance regarding Malthus' renewed formulation and he realized that it would become hard to eradicate the original, more popular, more insistent but wrong version:

> On the other side are those who maintain that population has a tendency (using the word tendency to express likelihood or probability) to increase beyond the means of subsistence; or, in other words, that, whatever be the existing means of subsistence, population is likely

[17] McCulloch (1863), part I, chapter VIII.

[18] As has been remarked, even the most recent attempts to resuscitate the classical Malthusian view seem to refer to a perception of history in which population would permanently and inevitably outgrow production as a "self-evident fact" and not as a "tendency".

> fully to come up to them, and even to struggle to pass beyond them, and is kept back principally by the vice and misery which that struggle must produce.[19]

Obviously, when confronting a person with this older Malthusian statement, he will most probably reasonably reject it and be inclined to consider this viewpoint as empirically falsified. If he, moreover, regards this argument as being the central one of the Malthusian theory of population, he will also erroneously be inclined to reject the principle of population. Consequently, when Mill and McCulloch employed phrases like

> that there is a *constant tendency* in all animated life to increase beyond the nourishment prepared for it, no one can possibly doubt[20],

Senior felt obliged to comment on their wording and clarified that

> we believe that they [Mr. Mill and Mr. McCulloch] have used it without being misled by it themselves, and, perhaps on that very account, without perceiving its tendency to mislead others. But that those whose acquaintance with Political Economy is superficial (and they form the great mass of even the educated classes) have been misled by the form in which the doctrine of population has been expressed appears to us undeniable. When such persons are told that "it is the tendency of the human race to increase faster than food." – "to people a country fully up to the means of subsistence", they infer that what has a tendency to happen is to be expected. Because additional population may bring poverty, they suppose that it necessarily will do so [...] [Such a doctrine] furnishes an easy escape from the trouble or expense implied by every project of improvement. "What use would it be," they ask, "to promote an extensive emigration? the whole vacuum would be immediately filled up by the necessary increase of population." [...] It is because we believe these misconceptions to be extensively prevalent that we have ventured to detain our readers by this long discussion. A discussion which some may think a mere dispute about the more convenient use of a word, and others an attempt to prove a self-evident fact.[21]

In summary, following Malthus' later editions, the principle of population will in this work be treated as a fixed "law" that continues to operate during the era of development (although potentially outweighed by some other effect), whereas his original statement of a Malthusian trap is considered as falsified after the breakout from stagnation. Consequently, since the end of the regime of stagnation does not mean the end of the principle of population, the positive effect of productivity growth on population must further be modeled as a part of the here advanced unified growth mechanism.

[19] Senior (1836), p. 146.
[20] Mill (1848), book I, chapter VII.
[21] Senior (1836), p. 149.

4.6 The Classical Theories of Population and Production in the Regime of Economic Development

4.6.1 An exogenous shock in fertility

Consequently, having shown that Malthus suggested that the principle of population displayed its full power to increase merely in theory, it should be asked by what forces its pressure has been attenuated, or "checked", in reality. Stating in his second proposition that

> population invariably increases where the means of subsistence increase, unless prevented by some very powerful and obvious checks,[22]

he implicitly determined the conditions under which population would not hit the limits of subsistence. Having exhaustively determined productivity as the ratio of production to population, the pressure of population could only be relaxed by either "increasing the means of subsistence" Y or by "checking powerfully and obviously" population N. To advance the latter remedy, we will first have to exhaustively define the checks to population, employing again the definition of a change in population, i.e. $\Delta N = B - D$.

> Mr. Malthus has divided the checks to population [N] into the preventive and the positive. The first are those which limit fecundity, the second those which decrease longevity. The first diminish the number of births [B], the second increase that of deaths [D]. And as fecundity and longevity are the only elements of the calculation, it is clear that Mr. Malthus's division is exhaustive.[23]

Hence, the three dinstinct remedies eligible for mitigating the pressure of population and consequently raising the level of productivity are "positive checks", "increasing the means of subsistence" and "preventive checks". It has been argued that neither positive checks nor increasing production are in the case of an unrestricted principle of population capable of releasing the pressure of population permanently and therefore raising productivity in the long run. When thus excluding these options as potential forces toward a more permanent increase in production per capita, it remains to evaluate the final option, i.e. to reduce the pressure of population by checking the number of births preventively and to conclude that the preventive checks are solely responsible for the "escape from the Malthusian trap". It is regularly overlooked that this result follows directly from one of Malthus' most crucial illustrations.

> In an endeavour to raise the proportion of the quantity of provisions to the number of consumers in any country [$y = Y/N$], our attention would naturally be first directed to the increase of the absolute quantity of provisions [Y]; but finding that, as fast as we did this, the number of consumers [N] more than kept pace with it, and that with all our exertions we were still as far as ever behind, we should be convinced, that our efforts directed only in this way would never succeed. It would appear to be setting the tortoise to catch the hare. Finding, therefore, that from the laws of nature we could not proportion the food [Y] to

[22] Malthus (1826), book I, chapter II.
[23] Senior (1836), p. 141.

> the population [N], our next attempt should naturally be, to proportion the population to the food. If we can persuade the hare to go to sleep, the tortoise may have some chance of overtaking her.[24]

Although often portrayed as a pessimist, Malthus saw the improvement of the individual economic situation as a very real possibility. Evidently, if population growth is restricted, the power of population will not be fully exerted. Moreover, if and only if the power of population is embanked, a situation is created in which production can possibly outrun population, generating per capita growth. Logically, apart from the positive checks, the only feasible way by which "the hare could be persuaded to go to sleep" was to propose birth control and hence to check population preventively.

> It is not in the nature of things that any permanent and general improvement in the condition of the poor can be effected without an increase in the preventive check; and unless this take place, either with or without our efforts, everything that is done for the poor must be temporary and partial. [...] This is a truth so important, and so little understood, that it can scarcely be too often insisted on.[25]

According to Malthus, the preventive checks include any action affecting the number of births, which reduces the maximum rate of fertility. Analogously to the case of the positive checks, he advised employing the level of the birth rate to measure the operation of the preventive checks, as this was the only way to exhaustively capture the preventive checks:

> The preventive check is perhaps best measured by the smallness of the proportion of yearly births to the whole population.[26]

Accordingly, wherever the preventive checks are strong, the birth rate will be observed to be low and vice versa.

Applying the foregoing considerations, it ought to be the decisive argumentation of a unified growth theory that only with the onset of the fertility decline population was increasingly prevented from eating up productivity gains, hence enabling economic growth in terms of GDP per capita. Moreover, this work intends to support the idea that the main distinction between the two regimes can be be reduced to their contrasting modes of population control. In particular, with the onset of the fertility decline, the era of stagnation – characterized by "positively checked" periodic overpopulation displayed by high death rates – is replaced with a modern growth regime, *because* a potentially abundant part of the population is constantly "preventively checked", i.e. via birth control, measured by low birth rates. We may therefore answer to the above mentioned skeptical economist: "Yes it's true – we observe increasing GDP per capita and at the same time decreasing population growth. But this does not contradict Malthus, it is part of his story. An increase in the preventive check *causes* population growth to decline which is in turn *an immediate cause for* the increase in productivity."

[24] Malthus (1826), book IV, chapter III.
[25] Malthus (1826), book IV, chapter XIII.
[26] Malthus (1826), book II, chapter XI.

4.6.2 An exogenous shock in fertility in the neoclassical model

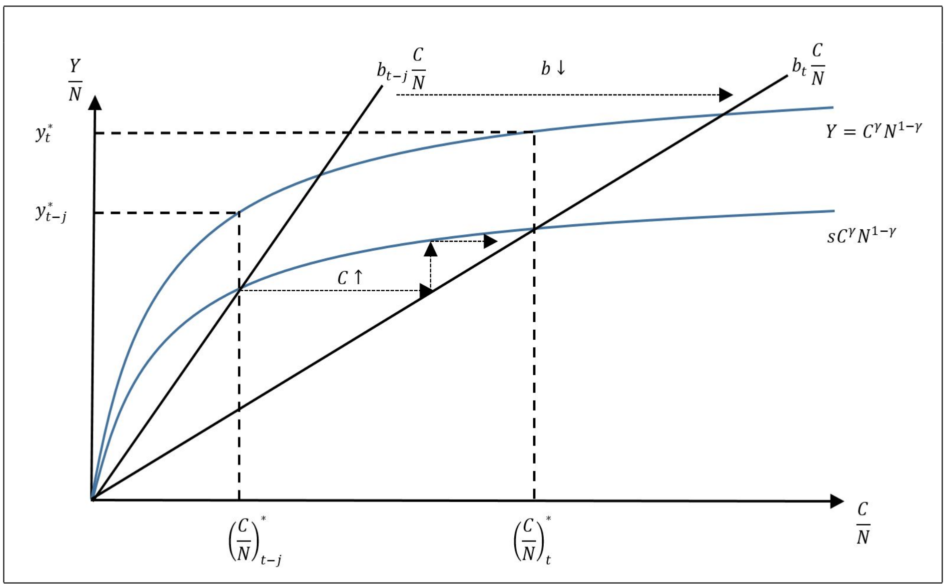
Using a neoclassical framework, this section illustrates mathematically and graphically that sustainably increasing productivity is predominantly the result of reducing too high fertility toward a lower level. This result is broadly in line with Galor (2011), who suspects that a substitution of child quality for child quantity backed the take–off toward a sustained path of economic development. In particular, Ashraf et al. (2013) find a negative effect of fertility on GDP per capita that can account for about 10% of long-run growth.[27] The classical framework here outlined argues that the historical reduction of fertility can almost completely explain economic long–run development.

It has commonly been argued that the neoclassical growth model used in chapter three is incomplete as it does apparently not account for the historically observed increases in productivity. As we have seen, this claim often overlooks the effect stemming from a potential decrease in the birth rate as was depicted in the structural equation of chapter three:

$$g_{y_{t,j}} = \frac{\gamma}{1-\gamma}(ln\ b_{t-j} - ln\ b_t) \tag{4.1}$$

While a higher birth rate reduces the steady–state value of productivity, we find that an exogenous *decrease* of the birth rate is well–qualified for causing productivity growth during the transition between two steady states. More explicitly, as is depicted in Figure 4.4, a continuously decreasing birth rate from b_{t-j} toward b_t is expected to yield continuous productivity growth, as in this case the right hand side of equation 4.1 will be positive.[28] The advanced theory suggests that economic development is, in this case, caused solely by the beneficial effects from the PoLD outweighing the detrimental effects of PoDR.

Figure 4.4: A population slowdown in the Solow model.



[27] The perhaps most recent evaluation of this argument is provided by Chatterjee and Vogl (2018).

[28] A result that has been confirmed by a number of studies on economic development including Sachs and Malaney (2002).

Furthermore, from the observed stylized facts it can be derived that population growth formerly seemed to outperform growth in production, causing stagnation, whereas in more recent times population growth is observed to have slowed down, offering the potential for economic development. Obviously, our neoclassical model fits perfectly into this framework, as it provides a well–established theory by which a decreasing birth rate is essential to allow productivity to increase. Consequently, the negative causal relationship from birth rate to productivity exhibited by the interrelation between PoDR and PoDL continues to operate over the whole time span under consideration and may provide the missing link between Galor's second and third stylized facts. To account for a more precise timing and magnitude of the relation between birth rate and GDP per capita growth, the parameters γ and j will be empirically estimated in chapter seven.

4.6.3 Simulation of a fertility decline

Recapitulating the above findings, the impact of a simple exogenous fertility decline on production and population can be simulated by introducing a negative trend l_t into the population equation of our simplified model of stagnation as follows:

$$\begin{aligned} g_{yt} &= \alpha_1 b_{t-15} - \alpha_2 b_t + \epsilon_{g_{y15}} \\ b_t &= \alpha_3 b_{t-1} + \alpha_4 g_{y_{t-1}} + \alpha_0 g_{l_t} \\ d_t &= \alpha_5 d_{t-1} \end{aligned} \tag{4.2}$$

Table 4.1: Calibration of the system of 4.2

α_1	α_2	α_3	α_4	α_5	α_0	b_0	d_0	g_{y0}	$\epsilon_{g_{y15}}$	l
see Table 3.1					0.4	see Table 3.1				0.020-0.001t

As is illustrated in Figure 4.5 and Figure 4.6, an exogenously modeled decline in the birth rate raises growth in production per capita by reducing population growth, due to – not inspite of – the continued operation of the three classical principles. Although productivity reacts with an exponential increase, the impact of the birth rate is not strong enough to account for the observed British sixteen–fold increase in GDP per capita, which is partly owed to the usage of the level of the birth rate instead of the logarithmized birth rate. Nevertheless, since we have so far neglected the effect of a decreasing death rate on population by holding mortality on a constant high level, the parallel decrease in production and population during the final periods of the simulation is not in line with the stylized facts. We will therefore deal with the mortality decline in the next chapter.

Figure 4.5: A simulation of an exogenous fertility decline in growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

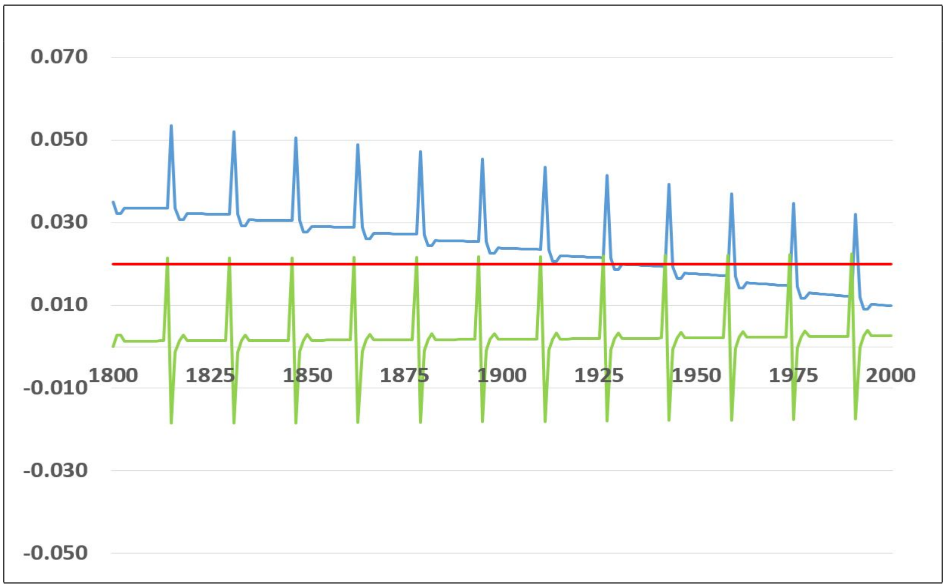
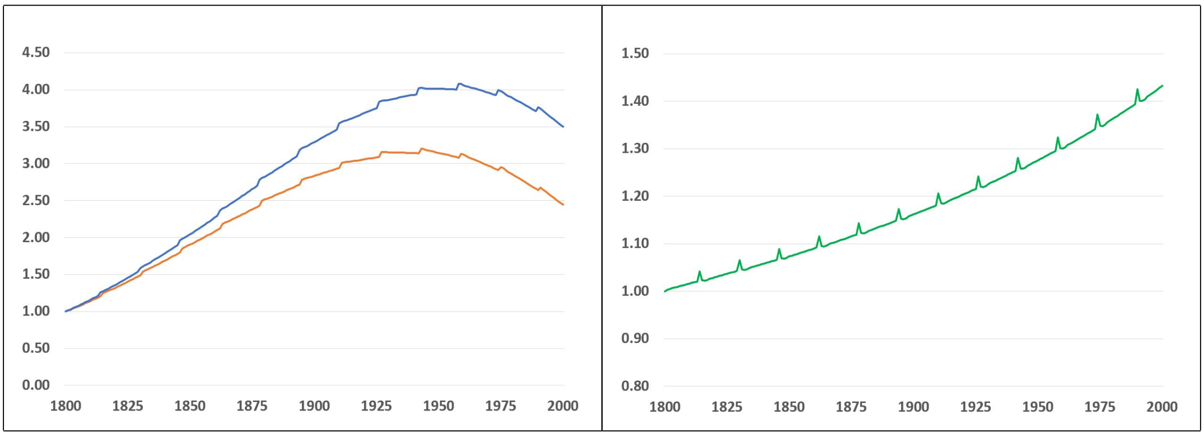


Figure 4.6: A simulation of an exogenous fertility decline in level variables (population (orange), production (blue) and GDP per capita (green)).



4.7 Fallacy 3: The Conventional Wisdom that "Development" Causes Fertility to Decline

So far, we have concluded that the fertility decline triggered the breakout from stagnation. Consequently, to understand the cause for this breakout we will have to answer the question of what induced fertility to decrease. Although we have thus shown that the fertility decline was more a cause for than a response to the increase in productivity, it is still very often unjustly reversely argued that the breakout from stagnation induced the birth rate to decline. This idea led to the notion that "development is the best contraceptive" instead of rightfully concluding that "contraception is the best development." Since we have modeled fertility to depend on productivity growth only in our simulation, this idea should at least be considered, as is done in this section.

4.7.1 An endogenous shock in fertility?

Based on the observed negative correlation between productivity and birth rate during the process of development, economic authors of the twentieth century began to believe that a rising living standard generally prompt individuals to lower their fertility — a currently quite popular idea that is widely known as the "demographic economic paradox".[29] A. Marshall[30] (1890) may already have prepared the way to mislead other economists by carelessly stating that

> on the whole it seems proved that the birth–rate is generally lower among the well–to–do than among those who make little expensive provision for the future of themselves and their families, and who live an active life: and that fecundity is diminished by luxurious habits of living.[31]

This quote might induce the reader to suspect a negative effect of high productivity (i.e. living standards) on fertility. As a consequence, nobel laureate G. Becker suggested that reaching a certain income or productivity threshold would increase the opportunity costs for having offspring and generally initiate a fertility decline.[32] The popularity of an income threshold might have partly been derived from the urge to find an economic justification for the Boserupian/Kremerian view after the original Malthusian view had (unjustly) been rejected, as Mokyr and Voth suggested that

> to avoid [population] showing explosive behavior [during the breakout from stagnation], a [fertility decline] is necessary, so that fertility responds negatively to higher incomes above some threshold level.[33]

There are, however, numerous arguments rejecting a potential negative impact of productivity on fertility. We will merely state four arguments here and return to this issue from an empirical

[29] See, for example, Becker and Lewis (1973).
[30] Alfred Marshall (1842-1924), British professor of political economy at Cambridge University, member of the Royal Commission in 1891, one of the founders of neoclassical economics.
[31] Marshall (1890) book IV, chapter IV.
[32] Becker (1981).
[33] Mokyr and Voth (2010), p. 9.

point of view in section 8.4. Firstly, since Becker's pioneering work, the existence of an income threshold could not be proven in spite of readily available data. Secondly, such a theory would imply that the formerly positive impact of productivity growth on fertility had not only mitigated but turned at once into a negative one. As Alter and Clark remarked,

> [e]conomic models of fertility face a fundamental challenge. All plausible models of population regulation for the pre–industrial world depend on a positive association between net fertility and income. This positive correlation of fertility and income became negative in Europe [during the nineteenth century], and there seems to be no association between income and fertility in high–income–low–fertility societies today.[34]

Thirdly, such a negative effect is the exact opposite and therefore incompatible with the evolutionary–proven principle of population. When looking at Sub-Sahara economies over the past sixty years, it seems more probable that this idea induced development policies unintentionally favoring population growth instead of what they were originally designed for, growth in productivity.[35] The general idea that "over feeding checks increase" had already been criticized by Spencer (1852) from an evolutionary viewpoint as follows:

> The theory which Mr. Doubleday [see also G. Becker] seeks to establish is, that throughout both the animal and vegetable – "Over feeding checks increase; whilst, on the other hand, a limited or deficient nutriment stimulates and adds to it." Or, as he elsewhere says, – "Be the range of the natural power to increase in any species what it may, the plethoric state invariably checks it, and the deplethoric state invariably develops it." [...] But how, under the alleged law, can a comparatively plethoric state ever be attained to? If the present production of necessaries of life is insufficient for the normal nutrition of the race, and if the resulting deplethoric state involves that the next generation will greatly exceed the present in numbers, then, for anything that appears to the contrary, the next generation will be in a more deplethoric state still. Unless Mr. Doubleday can show that the means of subsistence will increase more rapidly than the unduly fertile people, he cannot prove the existence of any remedial process. Nay, indeed, he must show that his law involves, under such circumstances, a greater increase of food than of people. Now he neither does nor can show this; and thus the alleged law lacks that very property of self-adjustment, which he rightly regards as the test of the real law.[36]

In other words, since it was shown in the last chapter that growth in production cannot outperform an unrestricted increase in population, the latter would create a generation even less productive, leading to a vicious cycle of higher fertility and lower productivity. Hence, the notion that producitivity growth would negatively affect fertility could never display an equilibrium as fertility would, in the long run, diverge to its maximum or minimum value. Fourthly, from an individual point of view, as will be described in chapter five, the existence of evolution by means of sexual selection proves that the choice of a partner strongly relies on its social rank, which is in turn in most cases determined by the level of income. Vice versa, the relationship offered here appears to contradict every model on fertility behavior, suggesting that sexual selection should

[34] Alter and Clark (2010), p. 63.
[35] Easterly (2001).
[36] Spencer (1852), Introduction.

fall on those potential partners exhibiting the greatest possible economic misery. In this case, idleness would be a permanently higher rated sign of sexual attraction than economic success, which is more than implausible from every point of view.

4.7.2 Human capital accumulation as preventive check?

More recently, Becker stated in his 1988 presidential address that economists should attempt to model the neoclassical growth model in one endogenous framework together with the "Malthusian growth model" to account for the escape from the Malthusian trap. Being closely related to an income threshold, he advanced a human capital view of fertility behaviour which will subsequently be briefly summarized, as it seems to represent the contemporary state of the art and to attract current research attention among unified growth authors – although this digression does not seem necessary for future investigations on the breakout from stagnation. Mokyr and Voth described the argument as follows.

> In many models of long–term growth, the fertility transition plays a crucial role, and the timing of fertility decline is central to many theories explaining the transition to self–sustaining growth. The decline is normally modelled as a response to changing economic incentives. Leading interpretations by Becker and Barro (1988) and Lucas (2002) emphasize the quantity–quality trade–off facing parents in a context of faster technological change and higher returns to human capital.[37]

Galor and Weil (2000), for instance, attribute an adapted fertility behavior to a sudden spurt in technology, raising the stock of knowledge.

> Eventually, parents invest more in the human capital of their offspring. This in turn accelerates the growth of knowledge. Higher incomes make it easier for parents to have more children. At the same time, a growing value of human capital produces incentives to increase the quality of one's offspring, reducing quantity. Initially, after the start of modern growth, the income effect dominated, leading to more births; later, the substitution effect became more important and fertility declined.[38]

However, Galor is mistaken in his belief that the increase in human capital as well as the increase in productivity preceded the decline in fertility, which is a logical conclusion given that he erroneously locates the beginning of the British fertility decline at the end of – instead of the beginning of – the nineteenth century.

Clark criticized both the theory of an income threshold as well as the theory of human capital accumulation by indirectly suggesting that the principle of population remained active after the breakout from stagnation:

> the evidence [...] is that those with the highest incomes and the greatest investments in the human capital of their children also had the largest numbers of surviving children.[39]

[37] Mokyr and Voth (2010), p. 10.
[38] ibid., p. 10.
[39] Alter and Clark (2010), p. 48.

Consequently, in this work human capital – as well as “technology” – is continued to be treated as a (endogenous) function of demographic variables, embodied by the principle of labor division, and is therefore omitted from consideration in our demographic unified growth theory.[40]

4.7.3 Other conventional factors as preventive checks

Lastly, Livi–Bacci (2012) lists a number of factors that have been suspected to have caused the fertility decline, including old–age insurance, culture, opportunity costs, emancipation, schooling, institutions, health and prohibition of child labor, most of which were certainly directly positively affected by concomitantly increasing productivity.[41] Nevertheless, this large variety of potential causes might confirm the rule that a long list of hypotheses merely reflects the helplessness of the scientific field involved. Hence, Mokyr and Voth concluded that

> finding an economic reason for fertility decline has not been easy, and there is currently no consensus on the principal contributing factors.[42]

[40] While “technology” tends to advance with the size of population, human capital tends to increase with longevity.
[41] Livi–Bacci (2012), p. 178.
[42] Mokyr and Voth (2010), p. 19.

Chapter 5

A Classical Growth Theory of Development

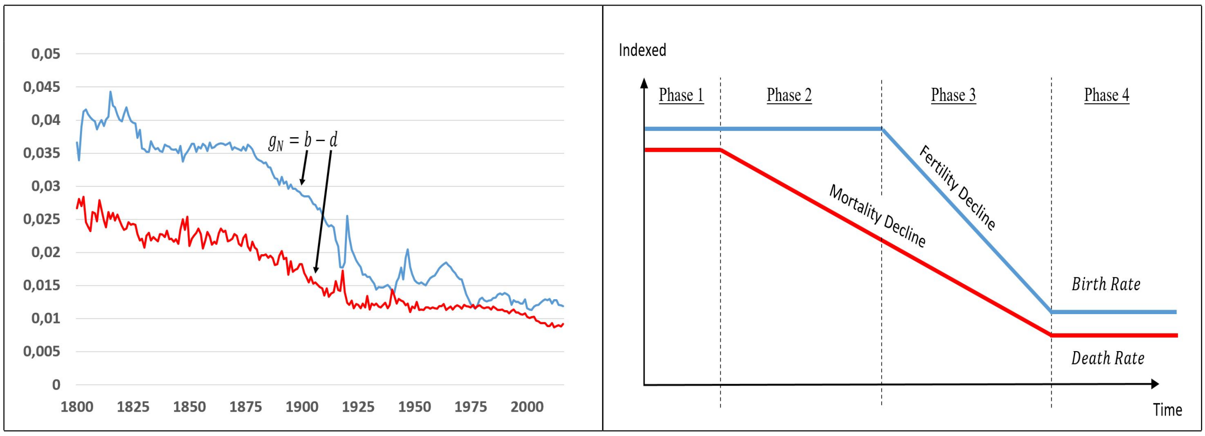
5.1 The Classical Theory of Population Part II: The Principle of Generation

Since our theory of production cannot explain the population slowdown endogenously, we will now return to the theory of population to examine the causes for the fertility decline. As they are crucial in elucidating modern economic development, this chapter again intends to resuscitate and to clarify the "vaguer intuitions" of the classical economists on the breakout from economic stagnation by providing the full classical version of the theory of population. Thus, this section suggests an explanation for the mechanism by which fertility was assumed to be regulated and by which the Malthusian trap can be avoided. Since we have seen that the observed increase in productivity does not appear to be able to explain the decline in fertility, we will now turn to the death rate as the remaining explanatory variable. As it constitutes the most abstract part of the theory, the great preventive check – the remedy to "escape the Malthusian trap" as suggested by the classical economists – deserves another elaborate treatment. Firstly, we will summarize the observed stylized facts regarding the death rate in relation to the birth rate before briefly exploring the existing theories by shortly reviewing the most recent literature. After having shown that this literature is still (or again) in its infancy, we will return to the classical authors and model their final proposition in accordance with the stylized facts.

5.1.1 The demographic transition

It has been shown that the population slowdown was the decisive factor in raising productivity. To analyze this population slowdown in detail, demographers usually decompose its evolution into a reduction of death rates, representing mortality, and a decrease in birth rates, representing fertility. The evolution of these vital rates over the past centuries is commonly conjointly stylized on a national level, illustrating a four–phase model termed as the "demographic transition" (see Figure 5.1). The model of the demographic transition in the right graph is crudely based on the British experience over the period 1800–2000 as is displayed in the left graph. In phase one, birth rates and death rates display relatively high values. At the beginning of phase two, the death rate declines due to the so–called "epidemiological transition," inducing as a by–product a population explosion, urbanization and sometimes an industrial revolution. In phase three, the birth rate starts to succeed the decreasing death rate, slowing down population growth, until they both settle on an approximately equal low level in phase four.

Figure 5.1: Left graph: British "demographic transition": Birth rate (blue) and death rate (red) 1800–2016. Right graph: Stylized fact "demographic transition".



Sources: Wrigley and Schofield (1981) for 1800–1871, Mitchell (2013) for 1871-2010.

Following the national accounts on the demographic transition, it must stare the neutral observer in the face that death rates and birth rates display some inherent connection. When comparing international data on vital rates, no economy can be found in which a sustained decline in birth rates has been realized without a preceding decline in death rates (see Figure 5.1 for Britain and again for numerous examples Appendix 11.2). For this reason, it is impossible not to become impressed with the idea that it appears to be a statistical law that the mortality reduction generally predates the fertility decline. Therefore, demographers have – in contrast to economists – ever since suspected a positive causal effect of changes in mortality on changes in fertility.[1] Although fertility seems to universally react belatedly to mortality, this lag varies strongly across economies, ranging from a few years to more than a century, which led many economists to wrongly shift their attention to the industrial revolution as main trigger for the fertility decline. Nonetheless, several economists also came to the conclusion that diminishing mortality must be causal for the fertility decline and the concomitant increase in GDP per capita.

The currently most cited economic explanation for a positive effect of mortality on fertility relies on the fact that decreasing mortality implies increasing life expectancy. Increased longevity may, for example, again foster human capital accumulation or decrease the time preference rate of individuals. Similar to the model by Galor (2011), where human capital accumulation was induced by technological progress, increasing life expectancy may be assumed to raise the demand for human capital and for child quality at the cost of child quantity.

[1] For a discussion see Kirk (1996).

> Cervellati and Sunde (2005) as well as de la Croix (2008) alter this [Galor's] setup by arguing that life expectancy rose quickly with productivity. This in turn encouraged investment in human capital, as payback horizons lengthened. Even if technological change is only slightly skill–biased, a self–reinforcing cycle of better technology, greater life expectancy, and higher investment in human capital can get started.[2]

Such models have been explored more recently by Ludwig and Vogel (2009), Herzer et al. (2012) or Cervellati and Sunde (2017b). As an additional benchmark, Cervellati and Sunde (2011) find a particularly strong positive correlation between mortality and fertility in those economies where life expectancy exceeds fifty years, suggesting a corresponding death rate threshold of about 0.02. The advantage of this group of models lies in their explicit integration of the demographic transition into a unified growth theory. However, as Mokyr and Voth criticize,

> models in the Lucas and Becker tradition emphasize the increasing demand for and returns to human capital, when we find little evidence of this [adding that] returns to human capital, conventionally measured, probably did not increase significantly before 1870.[3]

While the above association between mortality and fertility through the channel of human capital may be viewed as an indirect effect, there is growing agreement favouring a direct effect of mortality on fertility. A popular view of a direct mechanism emphasizes parental birth replacement behaviour when facing high child mortality. In addition to the well–known theory of infant replacement behaviour – to compensate for high infant mortality – the physiological effect of more rapid conception after early (infantile) deaths and the hoarding effect as an insurance against future high mortality when facing frequent subsistence crises are further direct effects on fertility.[4] Moreover, Clark even argues in favor of a rather "automatic" direct effect on the fertility decline due to a self–adjusting social environment as being the more promising approach:

> fertility limitation in northwestern Europe had little to do with rational individual calculation and much more to do with social customs. [...] Part of the evidence against conscious contraceptive practices is the lack of patterns in fertility that might be found where there was conscious control of fertility.[5]

As a result, direct effects of mortality might exist which do not rely on parents' conscious decisions. Social or religious habits such as the custom of late marriage may be capable of suppressing the potential number of children without conscious parental intervention, whereas a practice of divorces in liberal societies will tend to tear couples apart and impede excessive reproduction. Perhaps the hitherto best modern account of a potential mortality–fertility causation has been given by Hajnal (1965):

> If men had to wait till land became available, presumably a delay in the death of the holders of land resulting from declining death rates would tend to raise the age of marriage [and correspondingly diminish fertility] [...]; this is certainly a hypothesis that merits study.[6]

---

[2] Mokyr and Voth (2010), p. 11.
[3] Mokyr and Voth (2010), p. 41.
[4] See Kalemli-Ozcan (2002), Weil (2010) and Angeles (2010).
[5] Alter and Clark (2010). p. 47.
[6] Hajnal (1965), p. 133.

Being the authors main subject of research, it will subsequently be attempted to clarify the classical view on the great preventive check and to show that its operation has increased "without our efforts" as a response to the mortality decline.

5.1.2 The principle of generation as great preventive check

> At first sight it may appear that in any state of [a free] society [...] all the usual restraints to early marriage as they now exist would be removed, and that a rate of increase of the population unexampled in any previous era would be the result, leading in a few generations to a difficulty in obtaining subsistence, which Malthus has shown to be the inevitable result of the normal rate of increase of mankind when all the positive as well as the preventive checks are removed. As the positive checks – which may be briefly summarised as war, pestilence and famine – are supposed to be non–existent, what, it may be asked, are the preventive checks which are suggested as being capable of reducing the rate of increase within manageable limits?[7]

Problematically, Malthus (1803) seems to have inconsistently distinguished between the "preventive checks", the "preventive check" and the "great preventive check". He used the first two terms to define the "usual restraints" that were comprehensibly displayed by cultural traditions in most human societies. These traditions encompass those cultural customs explicitly and implicitly imposed such as a one-child policy, contraception, abortion, or linking the concession for legitimate marriage to the capacity to provide subsistence for a family. In contrast, the notion of the "great preventive check", which was betimes also abbreviated to the "preventive check", referred to a state of affairs in which these traditions were abandoned and individuals were left to their "natural and reasonable" decisions as a tool to restrict their fertility.[8] Although he argued that the great preventive check was crucial in preventing the population from growing exponentially, many classical economists were not able to follow his intuition and there seems to have been no definite agreement on the precise mechanism and definition of the "great preventive check" in classical economics.

Although Darwin (1859) had stated that high fertility was a dominant evolutionary strategy, Spencer (1874) suggested that the process of natural selection had in many species created a multitude of positive and preventive checks to avoid a permanent state of overpopulation. When looking at nature, it appears obvious that each species that has endured for millions of generations must, as soon as the available territory had been fully stocked and with the pace of specialization advancing very slowly, have exhibited a relatively stable population over this timespan. This, in turn, requires birth rates and death rates to be in equilibrium over the long run. Spencer proposed that the power of the positive checks was connected to the power of the preventive checks, writing that

---

7 Wallace (1890).

8 Malthus also employed the expressions "prudential restraint from marriage" and "moral restraint from marriage", since "marriage" was regarded to precede birth. Senior remarked on the use of the expression "marriage": "Our readers are of course aware that, by the word "marriage," we mean to express not the peculiar and permanent connection which alone, in a Christian Country, is entitled to that name, but any agreement between a man and woman to cohabit under circumstances likely to occasion the birth of progeny." Senior (1836), p. 143.

> proportioning of reproduction to mortality is requisite for mankind as for every other kind[9],

which he defined as

> the law of maintenance of all races; seeing that when they cease to conform to it they cease to be. [...] Individuation and reproduction are antagonistic.[10]

Indeed, fertility and mortality will be found to mutually balance each other. If fertility suddenly increased, a species must gradually become more numerous, until from lack of resources mortality would adjust to the level of fertility via the operation of the positive checks. If, conversely, mortality increased, then the species must diminish, until as a result of resources becoming relatively more abundant, fertility would rise to the level of mortality, as otherwise the species would become extinct. Also, it appears intelligible that a reduction of fertility eased the pressure on the means of subsistence and consequently might decrease mortality. However, the causal effects inducing fertility to adapt to diminished mortality, i.e. the preventive checks induced by decreasing mortality, are less clearly exposed. The mechanism of these most "natural" preventive checks will be illuminated in the following.

In an unchecked economy, unrestricted reproduction could be practiced by each individual as long as it was able to acquire the necessary resources. In this case, it has been argued that an excess of fertility had the tendency to ultimately force abundant individuals of the same generation into competition – a tendency that might generally be denoted as "intragenerational competition for niches". However, as has already been mentioned in chapter two, an excess of individuals and the concomitant pressure of population might, according to equation 3.8, alternatively emerge from a reduction in the death rate, diminishing the positive checks and mortality, increasing life expectancy and therefore raising the population share of older individuals. In this case, if two subsequent generations of individuals existed at the same time, a universally prolonged longevity would raise conflicts between the old, established and the young, emerging generation and correspondingly intensify "intergenerational competition for niches". The latter will be found to decisively cause the operation of the great preventive check by suppressing the fertility of the younger cohort.

5.1.3 The principle of generation in the animal economy: The struggle for territory and sexual selection

To inquire into the universal underlying causes that are responsible for a "conflict of generations", we may again first turn to the non-human economies. The strongest degree of intergenerational competition must be borne in the plant economy, where the possession of a natural niche almost exclusively relies on the availability of a fixed amount of territory. We may thus return to the initial statement that in a forest that is fully covered by beeches, it is impossible for seeds to start growing until an existing tree has died off. In this case, the conflict between subsequent generations itself constitutes the great preventive check in its most fundamental form, which

[9] Spencer (1874), §272.
[10] Spencer (1852), §2, §4.

we may name "the principle of generation" (in the following "PoG"). Among many bird and mammal species, where regular individual competition for territory is observed, procreation is likewise limited by the prevalence of an adult generation. In these instances, a relatively higher share of mature individuals established on a given amount of land tends to diminish the emerging generations' resources and timespan usually reserved for propagation and consequently the potential number of their offspring. This is most readily seen by observing the contrary fact that, if a "mortality crisis" induced the death of a large share of old, established individuals, a pool of newcomers would be readily available to take possession of the abandoned territory and strive to increase in numbers. As a result, intergenerational competition is raised by a decreasing death rate and relaxed by an increasing death rate.

What is more, in advanced species the great preventive check is quite considerably complicated by the existence of sexual reproduction. It is an important biological finding that among territorial species exhibiting different sexes, too high a fertility can be suppressed by a mechanism Darwin (1871) called "sexual selection". In cases where one sex is relatively abundant (in most species the males), the other sex can exert some choice over their potential partners. Among territorial species, possession of a territorial niche serves, beyond merely providing means of subsistence, as the decisive criterion of sexual attraction. Since, therefore, the possession of territory is an important condition for propagation, its occupation has evolved as the primary instinct of individuals of the abundant sex. The latter argumentation is based on observations made by J. S. Huxley[11]:

> Territory in some form or other is of prime biological importance in the life of birds (and probably of other groups as well). The first sign of sexual activity – the first effect, presumably, of the vernal change in the sexual organs – is in most species seen in the instinct of the males, not, as has usually been assumed to seek out the females, but to find, occupy, and defend a territory. So far as there is choice of mates in monogamous species, it is by the females, who seek out the males; but they only compete for those males who are in possession of territory.[12]

Given this form of sexual selection and that established individuals will already have acquired territory complementary attracting the other sex, nicheless individuals – in most cases young males – are regarded as unattractive and are therefore not considered for pairing, lowering the birth rate of the species. The operation of the preventive effect of sexual selection is more strongly exposed by restricting our attention, following Huxley, to monogamous species, where the attraction of one partner excludes the attraction of other potential candidates. Under this state of affairs, nicheless individuals – in most cases young females – are commonly not considered for reproduction and interbreeding is restricted to old, established pairs, further naturally reducing the reproductive capacity of the whole species.

Consequently, among monogamous territorial species, the great preventive check is, in addition to the usual degree of intergenerational competition, proportionally amplified by the degree of sexual selection. Under circumstances that concede low mortality, free choice of mating will

11 Sir Julian Sorell Huxley (1887–1975), British naturalist, biologist, first Director of UNESCO, founding member of the WWF, first President of the British Humanist Association, fellow of the Royal Society.

12 Huxley (1926), p. 148.

deny juvenescent male and female individuals the possibility to reproduce. This great preventive check is, as will be shown shortly, operating even more actively among the human species.

5.1.4 The principle of generation in the human economy: The struggle for a social rank, sexual selection and menopause

> But whence comes it, that the country where [...] the mean life, in whatever way the calculation is made, is higher than in any other, should be precisely that in which the fecundity is the smallest?[13]

One might be tempted to extend the fertility-preventing combined effect of intergenerational competition and sexual selection to the human case by returning to Malthus' view of a fully peopled pure pasture economy:

> Under such circumstances, how would it be possible for the young men who had reached the age of puberty, to leave their fathers' houses and marry, till an employment of herdsman, dairyman, or something of the kind, became vacant by death?[14]

Moreover, apart from the possession of a niche required for subsistence and sexual attraction, the decisive component leading to a drastic increase in the preventive effect in a regime of low mortality is due to the fact that human fertility is, particularly within monogamous couples, limited by age, preventively checking the potential fertility of old – in most cases female – individuals.[15] Malthus concluded that a strong degree of intergenerational competition would force an emerging generation to postpone reproduction until it will often be completely impeded by old age.

> The sons of farmers are exhorted not to marry, and generally find it necessary to comply with this advice, till they are settled in some business or farm, which may enable them to support a family. These events may not perhaps occur till they are far advanced in life. [...] Marriages would be among persons so far advanced in life, that most of the women would have ceased to bear children.[16]

In the following argumentation we will thus presume the existence of monogamy and a fertility interval limited by age in a human economy.[17]

The mindful reader will object that the suggested analogy projected from the bird economy to the human "territorial" economy masks an important Smithian characteristic of human societies, namely the existence of a social structure arising from regular exchange between individuals.[18] Since in human hunter and gatherer societies territory is in many cases not owned by single

13 M. Muret in Malthus (1826), book II, chapter V.

14 Malthus (1826), book II, chapter V.

15 This is again a simplification. It should be noted that menopause is not a purely human characteristic; see, for example, Ward et al. (2009).

16 Malthus (1826), book II, chapter VIII.

17 It should be noted that the change from the domestic institution polygamy to that of monogamy as well as the change from patriarchy to matriarchy are quite common and regularly observed phenomena among human as well as animal populations. See, for example, Spencer (1874).

18 It is obvious that the existence of a social structure is not solely restricted to the human species.

individuals, but by a community, scarcity of territory ceases to be the point of contention causing intergenerational conflicts. Correspondingly, sexual selection must in addition be exerted on other grounds than territorial ones. Nonetheless, it seems most plausible to assume that choice of mating will still tend to fall on those individuals that are assessed to be able to best provide subsistence for progeny. Indeed, in social economies, an individual's free choice appears to frequently center their attention on the social status, or as the classical economists called it, the "social rank" a potential partner appears to represent, which is quite reliably displayed by a corresponding "social niche", or in other words, a profession.[19] It is, hence, reasonable to replace the preventive effect resulting from the possession of a territorial niche with that resulting from the occupation of a social niche as a sign of attraction in human (social) economies.

Being thus confronted with a further criterion of sexual selection, the pursuit of territory must – from an evolutionary point of view – have been gradually complemented by a pursuit of social eminence as a "drive of prime biological importance." More explicitly, the average young individual must, under a strong degree of intergenerational competition, constantly strive to attain the former generation's social rank and consequently develop an instinct for social success, which is probably based on the experience of the parental success. Malthus, Senior and McCulloch suggested that the universal "fear of losing a social rank" would account for this additional instinct:

> Men will not be industrious without a motive; and the desire of bettering our condition, though powerful, is less so than the pressure of want, or the fear of falling to an inferior station. [...] With the lower classes the existence of present, and with the middle and upper classes the fear of future want, are the principal motives that stimulate intelligence and activity. The desire to maintain a family in respectability and comfort, or to advance their interests, makes the spring and summer of life be spent, even by the moderately wealthy, in laborious enterprises.[20]

Accordingly, while the pressure of want forced an individual of low rank to merely occupy *some* social niche, the fear of losing a social rank induced individuals exhibiting a higher social status to pursue those professions that retained their social position to impress the other sex with what Senior had called "decencies."

> The great preventive check is the fear of losing decencies, or, what is nearly the same, the hope to acquire, by the accumulation of longer celibacy, the means of purchasing the decencies which give a higher social rank.[21]

To illustrate the operation of the great preventive check, let us suppose a high-mortality-economy with a stationary population. Assuming the death rate to be 20 per thousand would correspond to a life expectancy of 50 years. Furthermore, suppose an inhabitant of this economy at the age of 25 whose parents - former physicians - have recently died at the age of 50, bequeathing their

[19] This positive relationship between income and marriage is in fact nothing but the microeconomic foundation of the principle of population, implying that pairing behavior in human and animal populations does not crucially differ.

[20] McCulloch (1863), part I, chapter VIII.

[21] Senior (1836), p. 144.

business to their child. Having acquired the parental social niche and consequently displaying the corresponding social status, the new physician will not hesitate to start a family. Now suppose mortality would fall over the next 25 years, such that the death rate was reduced to 12.5 per thousand, i.e. life expectancy would increase toward 80 years. In this case, the new physician's progeny is, at the age of 25, confronted with a new situation. Since their parents are well and alive, intergenerational competition arises, in most instances favouring the established generation. From the resulting inferior position, fearing the loss of the decencies they were used to growing up, the progeny will realize that they have to study medicine or experience additional medical on-the-job-training to be able to compete with the former generation to ultimately retain their social rank, until finally either the parental productivity has been achieved, or, as is much more common, the parents have retired or died. During the period of extended education, the offspring will generally neither commit to a partner, nor will they attract a potential partner of a corresponding social rank, thereby aggravating the finding–together of the sexes. Once the third generation has inherited the business and the social niche has been secured, it will again tend to propagate. However, assuming e.g. the parental retirement age to be 65 years, the newly established couple is most arguably too far advanced in life to produce their desired number of offspring, such that their potential fertility is correspondingly reduced.

Summarizing the operation of the great preventive check in the human economy, it might be stated that it is triggered by decreasing mortality and concomitantly intensified intergenerational competition for professions, preventing a young individual from occupying a social niche. The great preventive check comprises those actions stemming from "the fear of losing a social rank" that result in a postponement of reproduction until a later point in life. Accordingly, it must be remarked that its effect would be almost imperceptible if human fertility was not limited by age and is greatly reinforced by the prevalence of the domestic institutions monogamy and free choice of marriage. That the foregoing considerations with regard to the fourth classical principle, the principle of generation, are in accordance with Malthus' understanding of the great preventive check is highlighted by his most fundamental policy advice:

> I have stated expressly, that a decrease of mortality at all ages is what we ought chiefly to aim at. [...] It will be generally found true, that the increasing healthiness of a country will not only diminish the proportions of deaths, but the proportions of births and marriages.[22]

5.1.5 Direct and indirect effects of mortality on fertility

Following our above analysis, the last classical principle to be modeled refers to the great preventive check the "principle of generation", by which the power of population is repressed from peopling a country fully up to the limits of subsistence. Even Marshall apparently understood that the birth rate depended strongly on the availability of niches, writing that

> country life was, [. . .] rigid in its habits; young people found it difficult to establish themselves until some other married pair had passed from the scene and made a vacancy in their

[22] Malthus (1826), book V, chapter I and Malthus (1826), book III, chapter II.

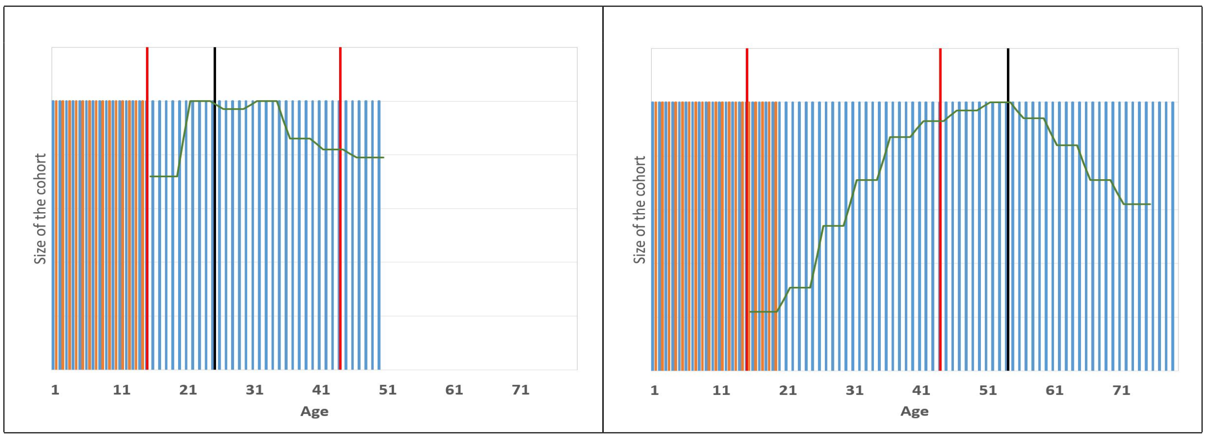
> own parish. [...] Consequently whenever plague or war or famine thinned the population, there were always many waiting to be married, who filled the vacant places.[23]

Accordingly, every individual faces the choice between reproduction and the preservation of its social rank during the early stages of its life. Further inquiries have shown that reproduction is in most cases not accomplished until a certain social rank has been achieved.[24] However, after a general increase in life expectancy (corresponding to the decline in mortality) and with the existence of a parent generation, a higher social rank cannot be achieved until the later part of life, postponing reproduction until the individual's average biological fertility interval has often been exceeded.

Nonetheless, a more precise mathematical formulation of the classical theory of population requires the principle of generation to be further analyzed to clearly distinguish between the particular effects of mortality on fertility. For, on the one hand, there exist mortality effects that directly act on fertility, notably an "inheritance effect" and an "infant mortality effect", while on the other hand, mortality effects operate indirectly through the income channel, weakening the effect of the principle of population. The latter will in this work be named "average income effect" and "selection effect".

To trace the origin of these effects, Figure 5.2 represents the stylized population structures for the years 1830 and 2010 respectively. For ease of illustration, populations are assumed to be stationary and stable, i.e. the birth rate equals the death rate and its relative age distribution does not change over time. The resulting cylindrical rather than pyramid form implies that every individual dies at the age of its life expectancy.[25] Average life expectancy can be recovered from the inverted death rate, which was roughly 2% in 1830 and 1.25% in 2010, excluding infant mortality.

Figure 5.2: Stylized population structure of England in 1830 (left) and 2010 (right). Displayed are working cohorts (blue shaded), cohorts in education (orange shaded), fertility intervals (red lines), average age of interitance (black lines) and average relative income (green lines) of a cohort.



Sources: Burnette (2006), U.S. Census Bureau (2011) for income statistics.

[23] Marshall (1890), book IV, chapter IV.
[24] See, for example, McCulloch (1863).
[25] The effect of early mortality is dealt with as part of the infant mortality effect.

To begin with the stylized population structure in 1830, individuals lived for fifty years on average, with the first fifteen years spent on "education." The fertility interval is taken to be constant, ranging from 15–45 years. As a result, 86% of the working population (benefiting from increases in income) in 1830 was fertile, whereas in 2010, when life expectancy was roughly 80 years, only 42% of the working couples were capable of reproduction (see blue shaded area). Accordingly, positive GDP per capita growth was in the latter situation increasingly distributed to infertile individuals of high age, who were not even physically able to convert the additional income into children. It is obvious that, if wealth is mainly distributed to an infertile population, Malthus' notion that "population invariably increases where the means of subsistence increase" largely ceases to be observable. This shift in social fertility is the first effect that can account for a breakout from the cycle of misery, the average income effect.

Furthermore, it can be observed that the life period during which the average individual earned its maximum income (green line) shifted from the young age of 20–35 years in 1830 to the old age of 45–60 years in 2010. As it is well known that individuals' choices of their partners are in a high degree positively affected by the latter's social rank, and as the individuals' social rank is quite reliably reflected by its relative level of income, it is a logical inference to presume a postponement of marriages between 1830 and 2010, resulting in an increasingly delayed fertility decision (selection effect).

Thirdly, the birth rate is directly affected by the death rate of those individuals who possess a part of the economy's wealth. With the death of such an individual, its property is usually bequeathed to the succeeding generation. Since the age of women at their first birth was approximately 25 years in 1830 and has not changed drastically over the last two hundred years and since their husbands are currently, quite similar to 1830, on average merely three years older, inheritance is quite universally passed to the offspring some 25–30 years before their own deaths.[26] Consequently, average age of inheritance was approximately 20–25 years in 1830 and around 50–55 years in 2010 (see black bar). Since early inheritance formerly allowed individuals to take over and make use of their parents' capital, often in form of a business, it tended to greatly increase their income and social rank, favoring "early marriage" and subsequently conversion of wealth into progeny. Until 2010 however, the channel for translating inherited wealth into a higher number of offspring was increasingly closed down, as the heir will, with a high probability, have arrived at an infertile age.

Complementing the above impact of the death of an old individual on fertility, the reduction of early deaths of individuals at a very young age completes the generation conflict by providing another well–known direct reason for low birth rates. When using a broader definition of the generation conflict, it also includes birth replacement behaviour. The diminution of infant and child mortality in the aftermath of the epidemiological transition seems to have induced parents to dispose of some formerly necessary replacement births.[27] Over time, this effect eased the social pressure on individuals to marry early, further postponing reproduction. For simplicity, the diminution of infant and child mortality will not be separately considered in this work, as

[26] See again Hajnal (1965) or Clark (2007) for historical marriage pattern.
[27] See, for example, Haines (1998).

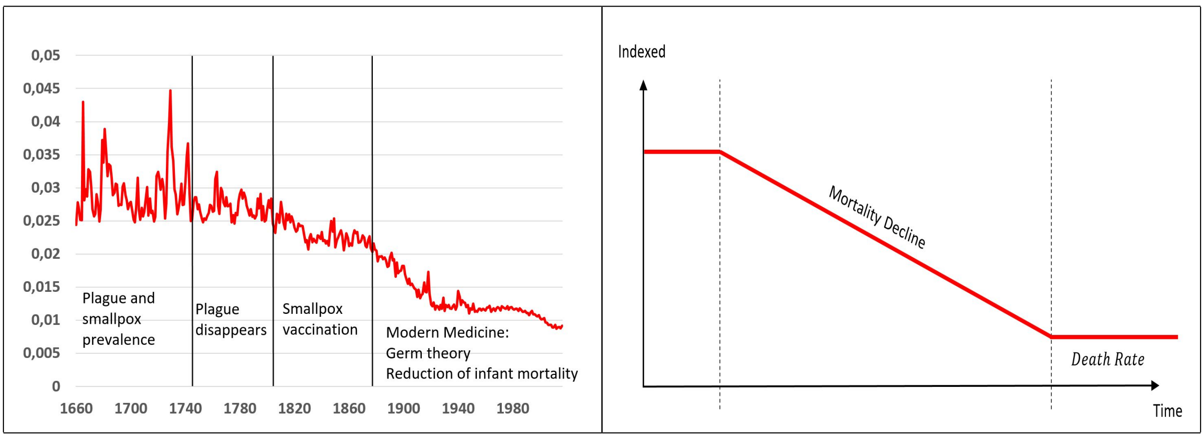
the effects of the eventual abolition of child replacements seem to be in line with the effect of the principle of generation as outlined above.

Summing up the outcome of these four effects of mortality on fertility, it might be stated that if two succeeding generations exist at the same time, a further rising life expectancy will progressively cause a generation conflict, forcing young individuals to preventively check their fertility. As a consequence, our mathematical framework will make use of the proposition that the birth rate is positively affected by the death rate. The operation of the principle of generation will be modeled by $\frac{\partial b_t}{\partial d_{t-x}} > 0$, again delayed by the cumulative lag of pregnancy and fertility decision. Moreover, we will find that, as in any animal economy, a reduction of the death rate below a critical threshold – in the human case 2% – will tend to be followed by an even stronger reduction of the birth rate. The corresponding human "take–off life expectancy" of about 50 years might be justified as being the starting value of a generational conflict, if it is defined as double the age of the average woman when having the first child, who will succeed in reproduction when adult. Social morals and customs are almost instantly readjusted to this conflict.

5.2 An Exogenous Trend in Mortality

Of course, if historically decreasing mortality was causal in reducing fertility, this begs the question of what triggered the death rate to decline. As has been hinted at, we find a general pattern of declining death rates among every developed economy often called the "epidemiological transition."[28] As is illustrated for the representative case of the British data displayed in Figure 5.3 over the period 1660–2010, the death rates seem to follow a continuous negative trend with initially high volatility abating over time. Next, we will shortly pinpoint the probable determinants of this mortality decline, i.e. the causes for the epidemiological transition.

Figure 5.3: Left graph: British "epidemiological transition": Death rate (red) 1660–2016. Right graph: Stylized fact "epidemiological transition".



Sources: Wrigley and Schofield (1981) for 1660–1871, Mitchell (2013) for 1871–2010.

[28] See, for example, McKeown (2009).

Research suggests the following two main factors as being responsible for the British mortality decline: The disappearance of the plague and the eradication of smallpox.[29] Although it seems difficult to find reliable British accounts of individual epidemics, the upward spikes during the early period are most arguably owed to regional plague epidemics, since no other cause of death could have claimed a comparably high number of victims and lists of plague epidemics in other European countries provide a similar pattern.[30] With the end of these epidemics, the spikes disappear, whereas the level of the death rate remained relatively unchanged between 1745 and 1800 until smallpox vaccination was introduced in the year 1798.[31]

> The wide swings in population size and real wages in Europe between the twelfth and the nineteenth centuries were primarily caused by external shocks, like the plague epidemics that reached Europe from Asia between 1240 and 1720. [...] After 1720 mortality fell as first plague was defeated and then smallpox, the latter through vaccination. [...] Consequently populations all across Europe expanded rapidly. [...] Vaccination campaigns dramatically reduced the incidence of smallpox after 1800.[32]

The latter allowed for a decrease in the level of the death rate over the subsequent thirty years, whereafter another 50 years of relative stability followed. It was only after 1875 that modern medicine emerged and seems to have enabled a further diminution of infectious diseases as well as a drastic reduction in infant mortality, which was largely completed by 1920.[33] Summarizing these observations, the British mortality decline until 1875 appears to be exclusively owed to the disappearance of plague and smallpox epidemics, hence the name of this shift, "epidemiological transition". While smallpox vaccination might be identified to have emerged from a "lucky medical coincidence", the end of plague epidemics has not been fully explained yet.[34] Naturally, economists have tended to ascribe the decrease in mortality to improvements in living standards in the form of better nutrition, ameliorating resistance to infectious diseases. If this was true, the demographic transition had to be modeled as a consequence of the rise in GDP per capita in form of better nutrition, opposing the demographic approach of this work. However, returning to Livi–Bacci's assessment,

> [o]ther, indirect, considerations also cast doubt on the nutritional hypothesis. For one, real wages in general declined throughout Europe during the eighteenth century and into the first decades of the nineteenth. [. . .] Another indication is variation in average height, which seems in this same period to have declined in England, in the Hapsburg Empire, and in Sweden. Height is fairly sensitive to changes in nutritional levels, and its decline or stagnation is certainly not a sign of nutritional improvement.[35]

29 For the former see Cipolla (1971), for the latter see Davenport et al. (2011)

30 "No one would deny that the disappearance of plague in the late seventeenth and early eighteenth centuries rid the Europeans of their most mortal enemy, and so reacted favorably on the development of population." Langer (1963), p. 4.

31 "Thereafter, the European history of epidemics — and indirectly also its population history — must be seen in a very different light. The subsequent demographic revolution would have been impossible in an environment that had been struck by the plague ten to fifteen times in every century." Cipolla (1971), p. 62 (translation by the author TL).

32 Alter and Clark (2010) p. 42, p. 56.

33 See Hays (2005) for a general overview on the history of epidemics.

34 The first intentional use of smallpox vaccination has most often been ascribed to Edward Jenner's method of inoculation of cowpox.

35 Livi–Bacci (2012), p. 72.

Instead, as Clark notes,

> [a] wide variety of explanations has been offered for the disappearance of plague, including viral mutations, competition between rat species, and new building materials.[36]

Furthermore, the traditional view that technological advances in medicine, for example in the form of hygiene or even quarantine triggered the epidemiological transition must be rejected as well:

> To answer this question [of the mortaliy decline] we must turn once again to evidence from the nineteenth century, when a decline in mortality was brought about by a reduction in the incidence of infectious disease, and was almost wholly independent of specific therapy. (The only medical procedure which can be accepted as having made a substantial contribution earlier than the twentieth century was vaccination, and its influence was limited to a single disease.) As stated previously, Griffith and others who considered this matter were completely mistaken in attaching great significance to growth of hospitals and other medical institutions.[37]

Consequently, in this preliminary, simple version of a classical growth model, the effects of higher productivity on the death rate will be put back, as wealth effects and technological advances seem to have played a minor role in the mortality decline.[38] Without dwelling too much on the nature of the decline and to not get lost in a long chain of causality, this work treats the disappearance of infectious diseases as an "exogenous" source for the epidemiological transition, unaffected by economic development and fertility. We may therefore adjust the third equation of our classical system by including a negative trend h:

$$d_t = \alpha_5 d_{t-1} - h$$

5.3 Simulation of Direct Mortality Effects on Fertility

As should have become clear from this chapter, in classical theory the great preventive check, the principle of generation, accounts for the missing link between the regime of stagnation and the regime of development and was formulated by Malthus to trigger the escape from the population trap. We will now again simulate the stylized facts on the breakout of stagnation. Firstly, in addition to the last simulation, we introduced a negative trend into the death rate equation. Secondly, to model the direct effects "inheritance" and "infant mortality" on fertility, the death rate will become part of the "population equation". We know from our law of population accumulation that $g_N = b - d$. In a fully stocked territory, where the population is not capable of improving its labor division and therefore $g_Y = 0$, long–run population growth is also zero, i.e. $b = d$, or written in growth rates

36 Alter and Clark (2010), p. 57.
37 McKeown (1965), p. 301.
38 We will briefly return to this issue in chapter seven.

$$ln(\frac{b_t}{b_{t-1}}) \approx g_{d_t} \Leftrightarrow ln\ b_t \approx ln\ b_{t-1} + g_{d_t}$$

By assuming that the birth rate is positively affected by changes in the death rate and with a delay (to account for the fertility decision), we add this effect stemming from the principle of generation to the effect of the principle of population depicted in the population equation. To put it differently, the growth rate of an exogenous trend in the birth rate simulated in chapter four is merely exchanged with the growth rate of an exogenous trend in the death rate as follows.[39]

$$\begin{aligned} g_{y_t} &= \alpha_1 b_{t-15} - \alpha_2 b_t + \epsilon_{g_{y_{15}}} \\ b_t &= \alpha_3 b_{t-1} + \alpha_6 d_{t-1} - \alpha_6 d_{t-2} + \alpha_4 g_{y_{t-1}} \\ d_t &= \alpha_5 d_{t-1} - h \end{aligned} \tag{5.1}$$

Table 5.1: Calibration of the system of 5.1

α_1	α_2	α_3	α_4	α_5	α_6	α_0	b_0	d_0	g_{y_0}	$\epsilon_{g_{y_{15}}}$	h
see Table 3.1					$\alpha_0 \cdot \frac{1}{d_{t-1}}$	0.4	see Table 3.1				0.0005

The calibration used for the following simulation is given by Table 5.1. As can be seen from the Table, the coefficient α_6 fluctuates with the death rate and might therefore be considered as "time–varying". For a better illustration, the logarithm of the birth rate is further omitted from this system of equations. As can be seen from Figure 5.4, due to the exact substitution of the linear exogenous trend in the birth rate with an exogenously diminishing death rate, birth rate and productivity stay unchanged when compared to the former simulation of chapter four. However, with the reduction of mortality, the simulation of the death rate is now in line with the stylized fact of the demographic transition, as can be seen from Figure 5.5. Nonetheless, population now appears to grow a little too fast compared with our real data series.

[39] The population equation can be derived from substituting g_{d_t} for g_{l_t}:$\alpha_0 g_{l_{t-1}} = \alpha_0 g_{d_{t-1}} = \alpha_0 \frac{d_{t-1}-d_{t-2}}{d_{t-1}} = \frac{\alpha_0}{d_{t-1}}(d_{t-1} - d_{t-2}) = \alpha_6(d_{t-1} - d_{t-2})$.

Figure 5.4: A simulation of an exogenous mortality decline in growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

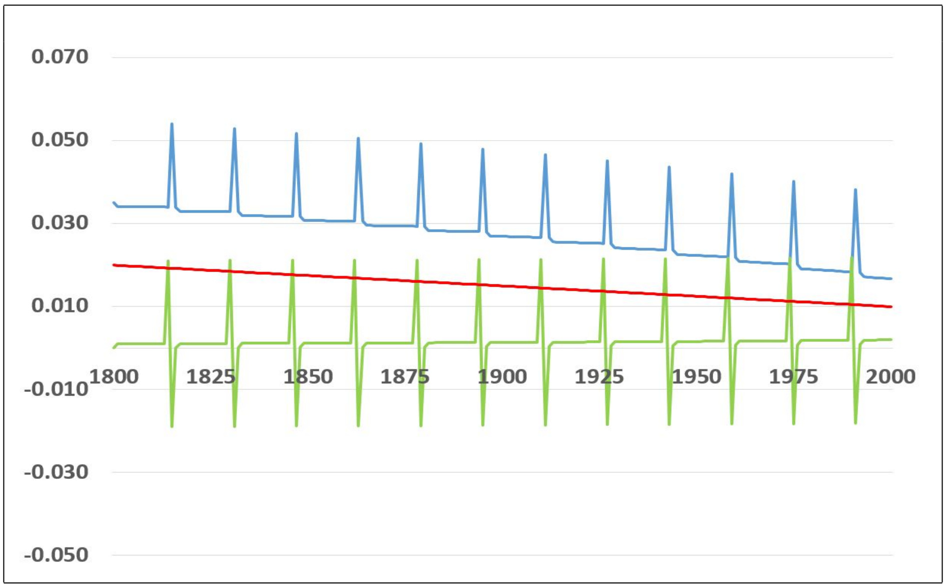
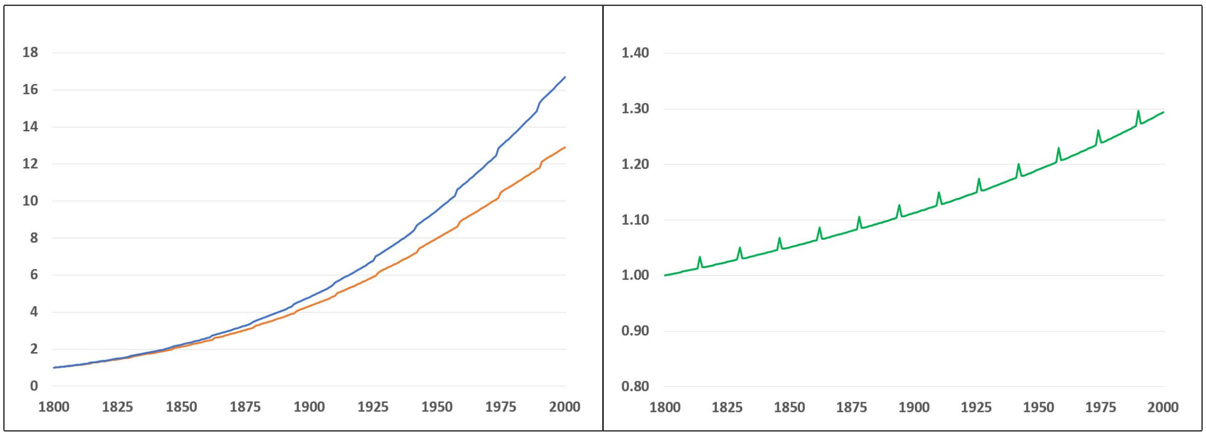


Figure 5.5: A simulation of an exogenous mortality decline in level variables (population (orange), production (blue) and GDP per capita (green)).



5.4 Simulation of Indirect Mortality Effects on Fertility

Finally, we have argued that mortality affects fertility indirectly through the income channels "average income effect" and "sexual selection effect". In this framework, these two effects are modeled to weaken the operation of the principle of population. If the principle of population had kept the birth rate on a high level for centuries and was responsible for stagnation, it must have been outweighed (see last section) or *suppressed* by some other principle during the nineteenth and twenthieth centuries. In analogy to the modeling of the direct effects, the coefficient α_4 will be tied to the death rate, reflecting the indirect effects, and might as well be considered to be "time–varying". In this case, however, the coefficient will be assumed to *shrink* as long as the death rate decreases. Consequently, the introduction of an exogenous trend in the death rate leads "over time" to an increase in the effect of the principle of generation and a decrease in the effect of the principle of population.

$$\begin{aligned} g_{y_t} &= \overbrace{\alpha_1 b_{t-15}}^{PoLD} \overbrace{-\alpha_2 b_t}^{PoDR} + \epsilon_{g_{y_{15}}} \\ b_t &= \alpha_3 b_{t-1} + \overbrace{\alpha_6 d_{t-1} - \alpha_6 d_{t-2} + \alpha_4}^{PoG} \overbrace{g_{y_{t-1}}}^{PoP} \\ d_t &= \alpha_5 d_{t-1} - h \end{aligned} \tag{5.2}$$

Table 5.2: Calibration of the system of 5.2

α_1	α_2	α_3	α_5	α_6	α_4	b_0	d_0	g_{y0}	$\epsilon_{g_{y_{15}}}$	h
see Table 5.1					$50 \cdot d_{t-1}$	see Table 5.1				

With the exception of the coefficient α_4, the calibration of the former system of equations remains unchanged (see Table 5.2). The simulation of the complete mechanism of development is displayed in Figures 5.6 and 5.7. First of all, the general tendency of all series is well in accordance with the observed stylized facts on economic development. Birth rate and death rate decrease continuously, production and population display a parallel increase and at the same time divergence and productivity growth increase, resulting in an exponential increase in productivity. Moreover, we will in the next chapter observe that the simulated declining volatility of growth in productivity is also in line with the empirical data. Consequently, our simulation of the direct and indirect effects of the great preventive check "principle of generation" appears empirically plausible and can account for the escape from the Malthusian trap. Whether the average real economy is indeed governed by these four classical economic principles will be more precisely evaluated within the econometric framework of chapter seven and chapter eight. An overview of the theoretical achievements of Part II of this work is provided by Figure 5.8.

Figure 5.6: An extended simulation of an exogenous mortality decline in growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

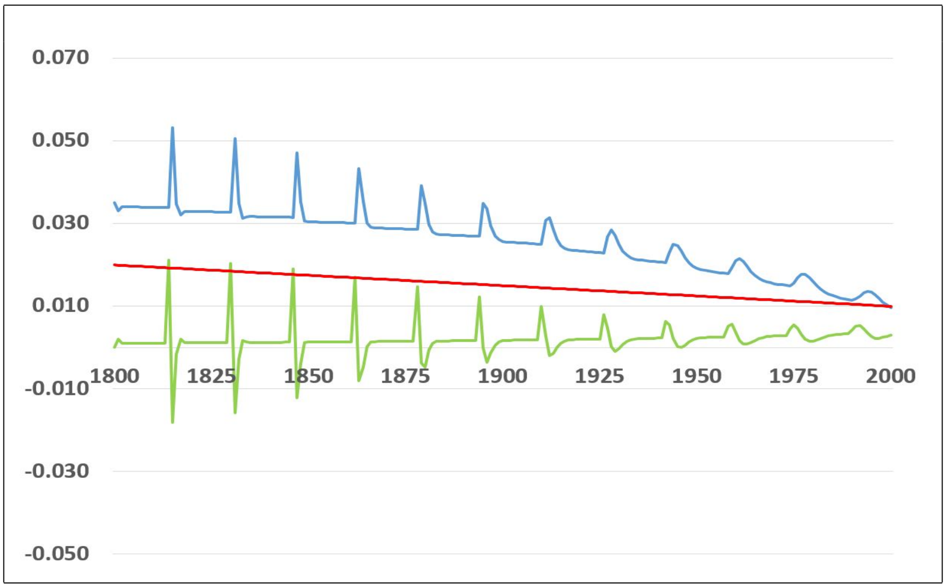


Figure 5.7: An extended simulation of an exogenous mortality decline in level variables (population (orange), production (blue) and GDP per capita (green)).

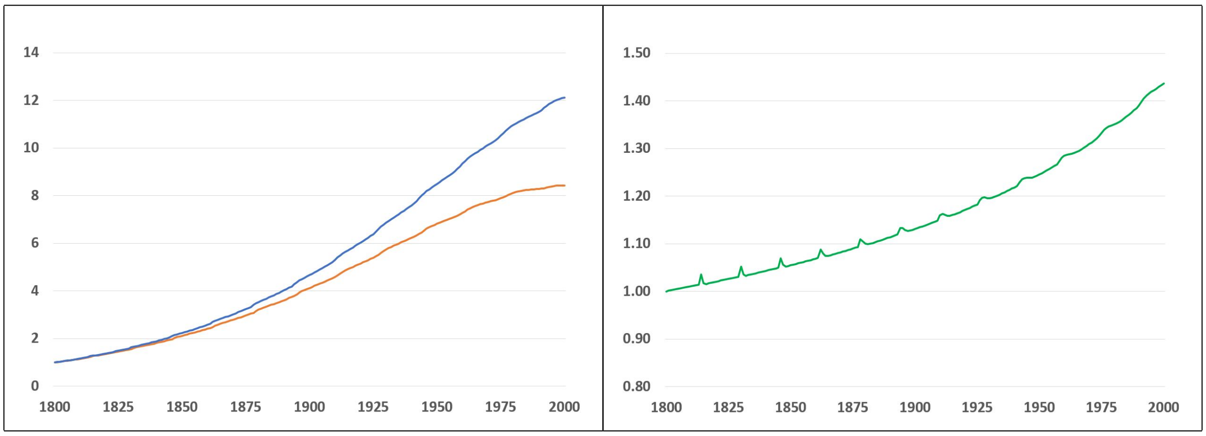
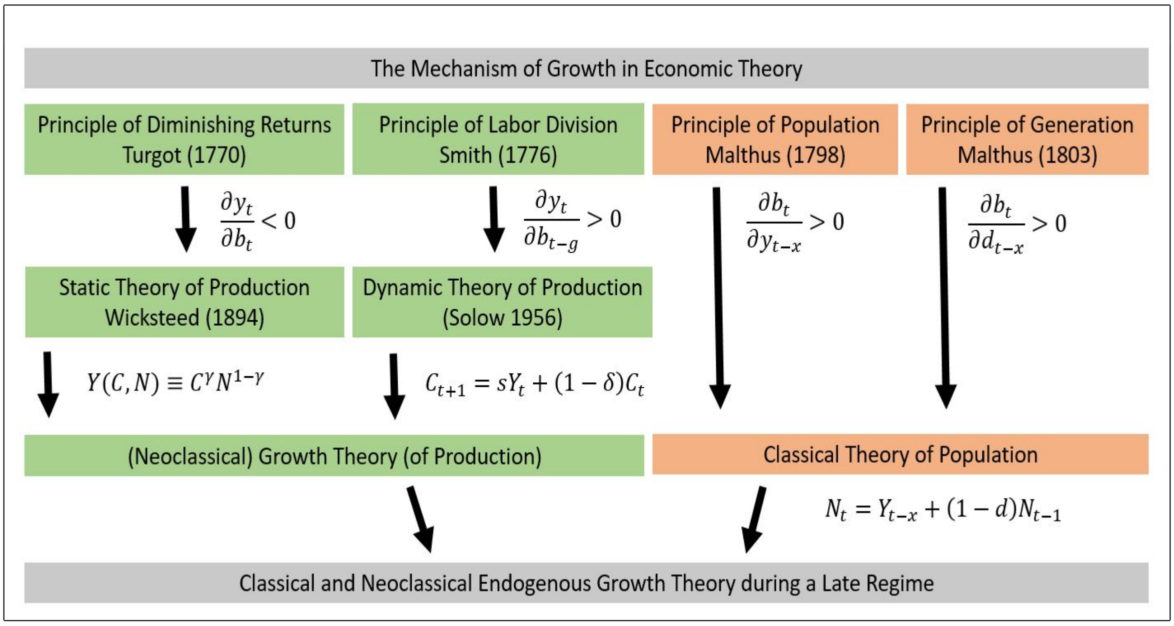


Figure 5.8: Theoretical findings of Part II.



Part III

Evaluation of the Classical Unified Growth Model

Chapter 6

A Classical Unified Growth Theory

> No plan for social improvement can be complete unless it embrace the means both of increasing the production of wealth and of preventing population from making a proportionate advance.[1]

6.1 Two Regimes of Stagnation and Development: The Stylized Facts Summarized

This chapter intends to unify the regime of economic stagnation and the regime of economic development in one economic growth model. From our descriptive analysis of economic and demographic variables from chapter two to chapter five, the subsequent set of stylized facts is viewed to be sustained by the British data displayed in the Figures 4.1 and 6.1, representing a universal global pattern of development.

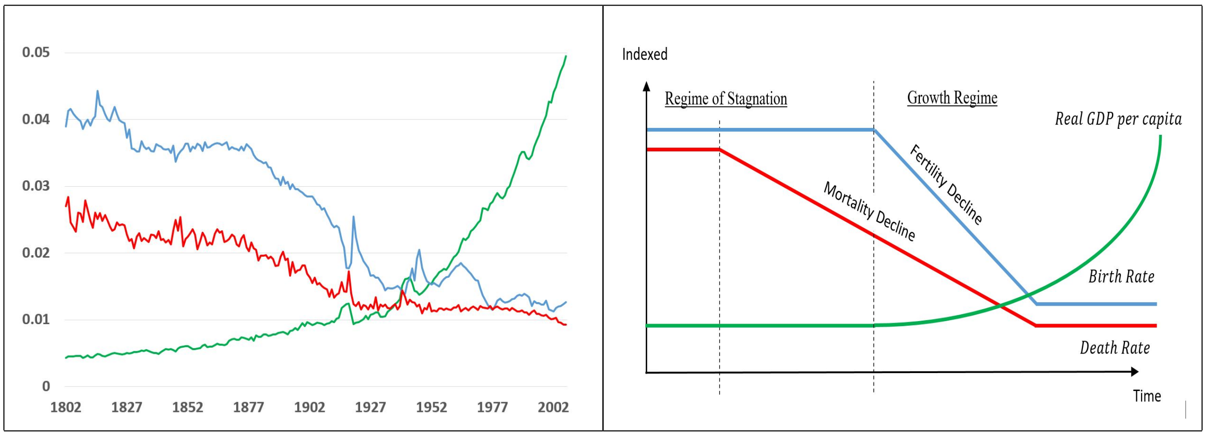
1. The cycle of misery: During the regime of economic stagnation, population grows proportionally with production (see chapters two and three).
2. The population slowdown: During the regime of economic development, population growth slows down and becomes inferior to growth in production. This slowdown is fully owed to a decreasing birth rate (see chapter four).
3. The cross of wealth: During the regime of development, there is some evidence of birth rates being negatively correlated with productivity: There is no modern economy in which productivity increased sustainably that has not gone through a demographic slowdown (see chapter four).
4. The demographic transition: During the transition to growth, there is some evidence of death rates being positively correlated with birth rates: There is no modern economy in which a decrease in the birth rate preceded a decrease in the death rate (see chapter five).

[1] Senior (1836), p. 146.

5. The epidemiological transition: The mortality decline was not initiated by a change in productivity, but — from an economic point of view — instead exogenously determined (see chapter five).

As is depicted in Figure 6.1, we conclude that the three growth variables birth rate, death rate and productivity growth follow this universal pattern of development. A decrease in death rates is categorically succeeded by a decrease in birth rates, causing the demographic slowdown, and a simultaneous rise in productivity growth. Once again, Appendix 11.2 provides numerous international examples in accordance with these stylized facts.

Figure 6.1: Left graph: British "transition to growth": Birth rate (blue), death rate (red) and GDP per capita (green) 1802–2007. Right graph: Stylized fact "transition to growth".



Sources: GDP: Clark (2009) for 1800–1871, Mitchell (2013) for 1871–2010, Vital Rates: Wrigley and Schofield (1981) for 1800–1871, Mitchell (2013) for 1871–2010.

To summarize the theoretical findings on the stylized facts, the following causal relationships are intended to constitute the cornerstones of a unified growth theory: Firstly, an unrestricted increase in population (due to a high birth rate) caused economic stagnation. Secondly, a population slowdown (due to a decreasing birth rate) was the crucial determinant to allow for a breakout from economic stagnation, i.e. economic development. Thirdly, the fertility decline is assumed to have been caused by the mortality decline. Finally, the process of economic development was ultimately triggered by the epidemiological transition.

6.2 Simulation of the Classical Unified Growth Theory:

When Senior (1836) published his treatise “An Outline of the Science of Political Economy”, he endeavored to summarize the collected scholarly principles of the time, or, in other words, the prevailing mainstream theory on economic growth. According to him, there existed common agreement among classical economists with regard to four elementary principles.[2] The definitions of the principles described in this work have been based on Senior's classification and have been named the principle of diminishing returns (PoDR), the principle of labor division (PoLD), the principle of population (PoP) and the principle of generation (PoG). These four principles have been integrated into an endogenous framework consisting of a theory of production and a theory of population.

The theory of production has been modeled by the first equation of the system in 6.1 displaying the effects of the birth rate on productivity growth via the PoLD and the PoDR. The theory of population has been modeled by the second equation of the system and exhibits the effects of both productivity growth and the death rate on the birth rate, representing the PoP and the PoG respectively. The last equation can be considered as link between an early regime of economic stagnation and a late regime of economic development and accounts for a unified growth theory. While an initially constantly high death rate allowed for the operation of the Malthusian trap mechanism, the switch toward a declining death rate induces the escape from the Malthusian trap as suggested by the classical economists. We may therefore properly term the theory advanced by this system a “classical unified growth theory”.

$$\begin{aligned} g_{y_t} &= \overbrace{\alpha_1 b_{t-15}}^{PoLD} \overbrace{-\alpha_2 b_t}^{PoDR} + \epsilon_{g_{y_{15}}} \\ b_t &= \alpha_3 b_{t-1} + \overbrace{\alpha_6 d_{t-1} - \alpha_6 d_{t-2} + \alpha_4}^{PoG} \overbrace{g_{y_{t-1}}}^{PoP} \\ d_t &= \alpha_5 d_{t-1} - hI \end{aligned} \tag{6.1}$$

$$with\ I = 0\ for\ t = 1, ..., 250\ and\ I = 1\ for\ t = 251, ...350$$

[2] “[Presuming] that every man desires to obtain additional wealth with as little sacrifice as possible:

1. That agricultural skill remaining the same, additional labour employed on the land within a given district produces in general a less proportionate return, or, in other words, that though, with every increase of the labour bestowed, the aggregate return is increased, the increase of the return is not in proportion to the increase of the labour. [PoDR]
2. That the powers of labour, and of the other instruments which produce wealth, may be indefinitely increased by using their products as the means of further production. [PoLD]
3. That the population of the world, or, in other words, the number of persons inhabiting it, is limited only by moral or physical evil, [PoP]
4. or by fear of a deficiency of those articles of wealth which the habits of the individuals of each class of its inhabitants lead them to require.” [PoG]

Senior (1836), p. 139.

Table 6.1: Calibration of the system of 3.13

α_1	α_2	α_3	α_4	α_5	α_6	b_0	d_0	g_{y0}	$\epsilon_{g_{y15}}$	h
1	1	1	$50 \cdot d_{t-1}$	1	$0.4 \cdot \frac{1}{d_{t-1}}$	0.035	0.020	0.000	0.025	0.0005

For calibration, coefficients and initial values from the former section are retained and the third equation is supplemented by an indicator function. The results from the simulation are displayed in Figure 6.2. The first 250 periods of the simulation correspond to the evolution of the regime of stagnation as it was modeled in form of the Malthusian trap, following a shock in $g_{y_{15}}$. The second part of the simulation ranging over the last 100 periods accounts for the regime of development and is triggered by the decline in death rates. This decline decisively induces the progressive operation of the PoG according to the second equation of 6.1. Owing to the direct mortality effects, the birth rate eventually declines even more rapidly than the death rate. In the case of the indirect effects, the short–run conversion of productivity into births owing to the PoP decreases in magnitude. The potential for economic growth is triggered by the fact that birth cohort size decreases over time. If the term $\alpha_2 b_t$ was larger than $\alpha_1 b_{t-15}$, the negative effect of the PoDR due to an ever–growing population outweighed the positive long–run effect owed to the PoLD. However, as long as the ratio $\alpha_1 b_{t-15} > \alpha_2 b_t$, i.e. the birth rate declines over the course of one generation as is observed in Figure 6.2 after period 250, the ratio between unproductive and productive individuals abates as well. In this case, the productivity gains from labor division outperform the losses from diminishing returns. This simulation affords a confirmation of the suspected regimes of stagnation and development matching the stylized facts, furnishing the classical (unified) growth theory with a consistent mathematical framework. In the following two chapters, we will evaluate in how far this system of equations is confirmed empirically. To this end, we will estimate the short–run relationships suggested by the theory of population using a vectorautoregression on the complete system 6.1 in chapter seven. Eventually, the long–run relationship between production growth and birth rate will be estimated in chapter eight by employing simple OLS–regressions to the first equation of system 6.1 only.

Figure 6.2: A simulation of classical unified growth theory in growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

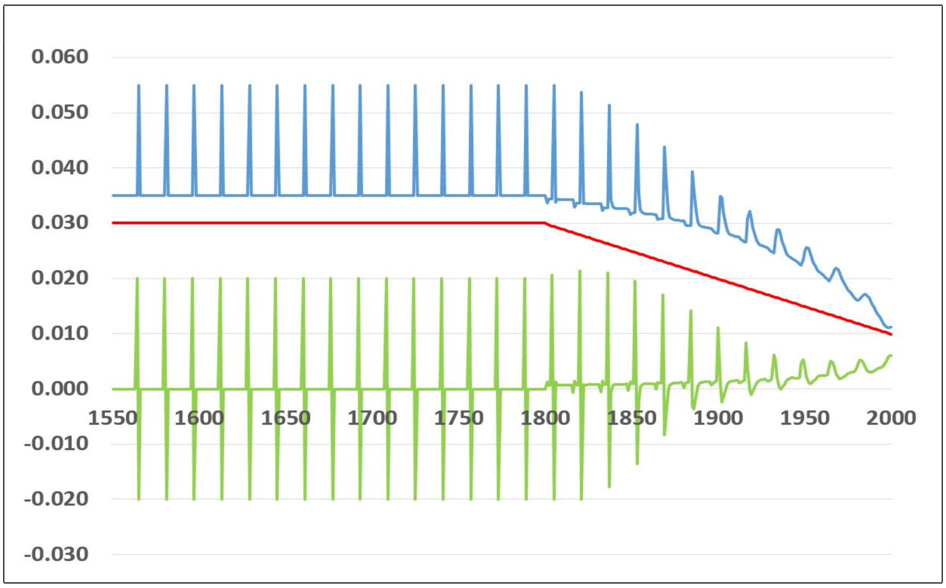
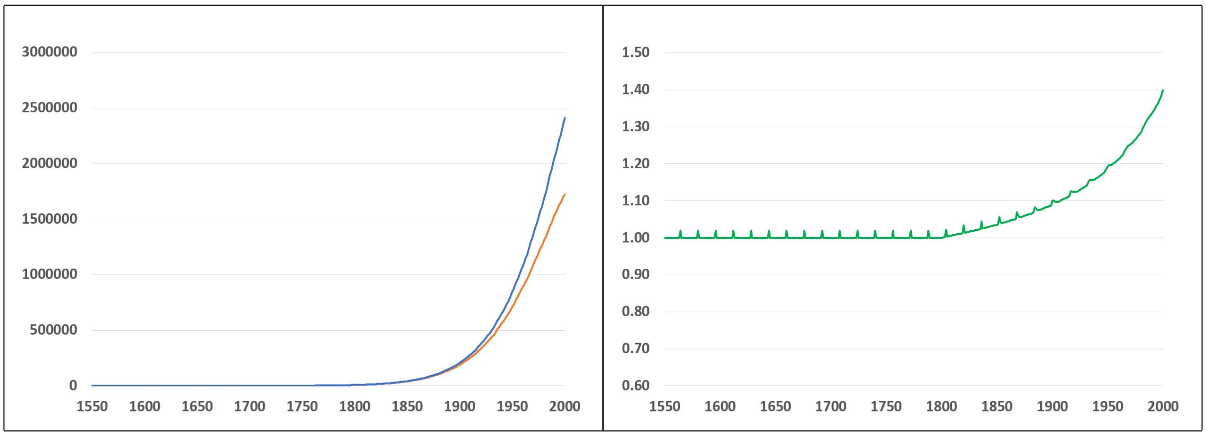


Figure 6.3: A simulation of classical unified growth theory in level variables (population (orange), production (blue) and GDP per capita (green)).



Chapter 7

Measuring the Classical Theory of Population

7.1 Methodology

7.1.1 Data

Since the emergence of classical political economy, economists have tried to make sense of the apparent link between demographic and economic variables observed in the stylized facts. In attempting to verify the four classical principles empirically, statisticans face the difficulties of deficient preindustrial data and endogeneity between demographic and economic variables. Regarding the data, with the construction of Wrigley and Schofield's (1981) preindustrial time series on birth rates and death rates, quantitative studies were able to yield evidence of falsifiable hypotheses for the British case, although Clark's data used here on preindustrial GDP per capita are – as we have seen – still debated. Moreover, in search of an overarching unified growth theory formed by universal principles, it is essential to compare the existence of the principles internationally. Mitchell's international historical statistics arguably provide the longest and most comprehensive official national series on vital rates and GDP per capita.[1] This historical database, partly corrected by the author to eliminate some obvious typing errors, will be made use of in this chapter.[2]

7.1.2 Vectorautoregression

For dealing with endogeneity, the author regards time series analysis as being the most appropriate tool. Lee (1981) was the first author to evaluate the relationship between vital rates and economic variables by employing distributed lag regressions on Wrigley and Schofield's dataset. Eckstein et al. (1986) were one the first to test economic hypotheses using a VAR model. Using

[1] Since population is an internationally relatively immobile factor of production, an estimation on the country level is expected to yield significant results.

[2] The adjusted data can be found on the attached data media.

a Vector–autoregressive (VAR) model allows for solving the problem of endogeneity by treating all variables as endogenous.[3] However, a VAR model must be applied with care, as its form strongly depends on the underlying theoretical foundation. Nicolini (2007) refined Lee's approach by using a VAR model and illustrating impulse response functions that allowed assessment of the qualitative relationships between the three variables. Building on the VAR and developing a more complex methodology, Herzer et al. (2012) employed a Vector error correction (VEC) model to account for possible cointegration between the variables, while Rathke and Sarferaz (2014) introduced time-varying coefficients.

This paper will retain the traditional VAR approach, since it is most arguably the easiest and most transparent estimation, as it mainly requires the knowledge of simple OLS estimation. The idea of the VAR approach is to recover the relevant coefficients from an OLS regression of contemporary values on lagged values of the variables and to use the recorded parameters to project the average impact of an exogenous shock in one of the variables over time. The obtained impulse response functions are expected to conform to the classical principles as formulated in chapter three. However, for the linear system to qualify as a VAR representation, some further reservations will be made in the following. To evaluate the principles in question, the system of our unified growth model constructed in the last chapter might initially be written in matrix notation as

$$\begin{pmatrix} gy_t \\ b_t \\ d_t \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ \alpha_4 & \alpha_3 & \alpha_6 \\ 0 & 0 & \alpha_5 \end{pmatrix} \begin{pmatrix} gy_{t-1} \\ b_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\alpha_6 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} gy_{t-2} \\ b_{t-2} \\ d_{t-2} \end{pmatrix} + \begin{pmatrix} 0 & \alpha_1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} gy_{t-15} \\ b_{t-15} \\ d_{t-15} \end{pmatrix} + \begin{pmatrix} 0 & -\alpha_2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} gy_t \\ b_t \\ d_t \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -h \end{pmatrix}$$

To apply the VAR approach consistently, the presumed real relationships are expected to be *linear*. While the estimations of Lee, Nicolini, Herzer et al. and Rathke and Sarferaz were usually based on the usage of a level variable of real wages or real GDP per capita, they will in this case be replaced with growth rates of real GDP per capita, since the major part of the true relationships between the variables of our system becomes linear only when employing growth rates as has been justified in the theoretical parts of his work. Moreover, as growth rates display the same internationally valid unit of measurement, the principles may be simultaneously estimated across countries. Thirdly, instead of level variables, growth rates are most arguably stationary, which is required to avoid spurious autoregressions when not specifically accounting for cointegration.

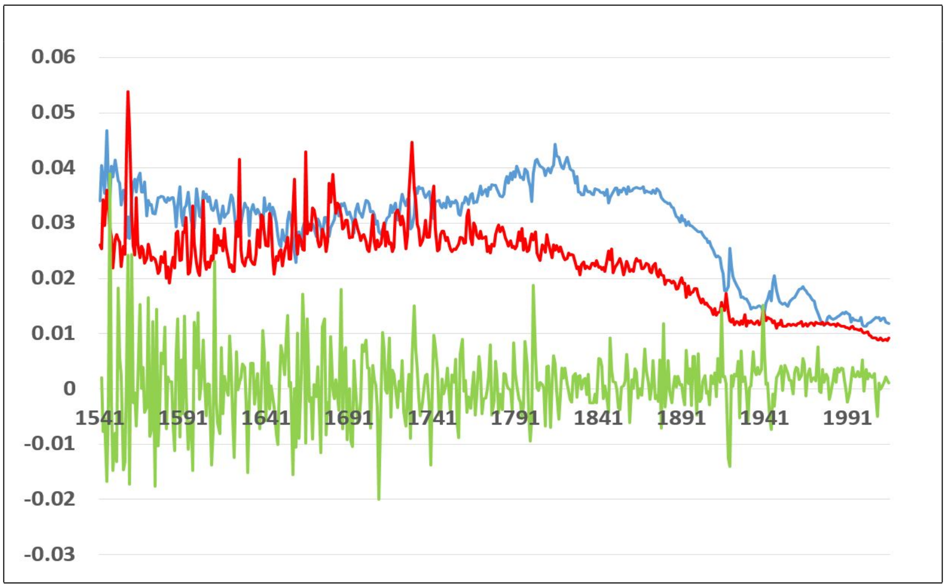
7.1.3 Stationarity of the variables

An OLS estimation over time requires at least some of the data series to be stationary, as integrated or trended variables will almost certainly give spurious results. Since the British data provide the longest national time series available, ranging from the year 1541 to 2010, tests on the order of integration as well as the tests for lag selection will be representatively conducted on this sample. The annual data on which the VAR model will be based are displayed in Figure 7.2.[4]

[3] See, for example, Lee and Anderson (2002) and Crafts and Mills (2009).

[4] GDP per capita growth is divided by ten for better visualization.

Figure 7.1: Britain: Time series on birth rate, death rate, GDP per capita growth 1541–2010. Sources: Clark (2009), Mitchell (2013), Wrigley and Schofield (1981).



Sources: GDP: Clark (2009) for 1541–1871, Mitchell (2013) for 1871–2010, Vital Rates: Wrigley and Schofield (1981) for 1541–1871, Mitchell (2013) for 1871–2010.

In the case of GDP per capita growth, the results from running Augmented Dickey–Fuller tests on non–stationarity seem to unequivocally indicate stationarity of the variable (see Table 7.1), while the application of the same test to death rate and particularly to birth rate does not always reject the null hypothesis of non–stationarity on a 1% level (see Table 7.1). Indeed, the pattern of birth rate and death rate has led to a debate on their order of integration.

Table 7.1: Unit root tests on the relevant variables.

Augmented Dickey–Fuller test for unit root			Number of obs = 468		
model	1% Crit. Value	Test Stat. GDP pc gr	Test Stat. Birth Rate	Test Stat. Death Rate	Test Stat. Pop growth
2 lags, no constant	-2.580	-16.737***	-1.235	-1.451	-3.747***
2 lags, constant	-3.443	-17.460***	-0.941	-3.539***	-6.303***
2 lags, linear trend	-3.981	-17.620***	-1.682	-5.925***	-6.304***

*** indicates significance at 1% level

Firstly, following Nicolini (2007), vital rates could be treated as stationary variables, as their values represent (population) growth rates and are by definition restricted to lie within the range (0,1). Besides, despite vital rates displaying high persistence, it seems implausible to believe that they have ever exceeded a certain maximum value, say ten percent, or that they have fallen below a minimum value, say zero, in the long run. Accordingly, they cannot in reality follow a random walk or a trend and the assumption $0 < \alpha_1, \alpha_3 < 1$ should hold. Nevertheless, stationarity of these two variables might be questioned by having found evidence of the variable natural population growth rate being stationary on a one percent level. As the latter is by definition a linear combination of birth rate and death rate, there is strong indication for the vital rates being cointegrated.[5] However, as was pointed out by Fanchon and Wendel,

> "VAR models can be estimated with data on stationary and non-stationary variables if the non-stationary data is also cointegrated because recent theoretical work proves that estimation with such data will yield consistent parameter estimates."[6]

Thirdly, as was suggested by Sims (1980), if we are more interested in the nature of relationships between variables with the end purpose being estimation of the impulse response functions to capture the dynamic responses and less interested in point estimates, estimating a VAR with non-stationary variables can give us important insights into their relationships. Accordingly, the question of the order of integration of the vital rates does not pose problems with regard to a consistent estimation of a VAR model.

7.1.4 Lag order selection

In the foregoing simulation, the benefits from the division of labor were strongly simplified. However, there are at least two important reasons complicating their measurement in empirical analyses. Firstly, since we will employ national data without accounting for an international labor division, the effects of foreign population growth on domestic output are not captured in the regression. Since external trade shocks might be suspected to cause a major part of the strong fluctuations of GDP per capita data as shown in Figure 7.1, this effect should not be underestimated. Secondly, to roughly illustrate the positive delayed effect of births on the extent of the division of labor, a lag of fifteen years was employed in the simulation. For all real applications, the exact timing of an average individual entering the division of labor cannot be sufficiently determined, much less the resulting benefits, which are arguably spread over the subsequent lifetime. It is assumed that a VAR model is too "costly" in terms of parameters to be able to properly estimate the effect of the PoLD, which is why the fifteenth lag will be eliminated from estimation. This issue will be dealt with in chapter eight. Vice versa, omission of the fifteenth lag greatly increases the number of degrees of freedom, which is particularly valuable when using small sample sizes.

Nevertheless, it is advisable to include a third lag by which the additional information stemming from the PoLD, stored in the remaining error terms, might be captured. The use of a VAR(3) model is supported by running a series of lag–selection tests on the British data, as the most parsimonious model is suggested by the Schwarz–Bayesian information criterion to use three lags (see Table C1 in Appendix 11.3). With regard to a delayed fertility decision when accounting for the PoG and the PoP, a lag of three years appears plausible as well, whereas every additional lag

[5] Using a VEC model specification similar to that of Herzer et al. (2012), explicitly accounting for the cointegrated variables birth rate and death rate, or estimating a restricted structural VAR model do not yield very different results.

[6] Fanchon and Wendel (1992).

may unnecessarily increase the number of parameters to estimate. Replacement of the fifteenth by the third lag gives

$$\begin{pmatrix} gy_t \\ b_t \\ d_t \end{pmatrix} = \Phi_1 \begin{pmatrix} gy_{t-1} \\ b_{t-1} \\ d_{t-1} \end{pmatrix} + \Phi_2 \begin{pmatrix} gy_{t-2} \\ b_{t-2} \\ d_{t-2} \end{pmatrix} + \Phi_3 \begin{pmatrix} gy_{t-3} \\ b_{t-3} \\ d_{t-3} \end{pmatrix} + \Phi_0 \begin{pmatrix} gy_t \\ b_t \\ d_t \end{pmatrix},$$

where the trend h employed in the simulation has been excluded, since it is supposed to be empirically captured by Φ_1.

7.1.5 Ordering of the variables

The estimation of an unrestricted VAR model of the above system of equations is complicated by its inclusion of contemporaneous effects required to measure the PoDR and other possible immediate effects. To analyze the interactions between annual demographic and economic variables, Nicolini (2007) proposed a recursive VAR structure based on Theil (1971) of the vector form

$$A_0 Y_t = \sum_{j=1}^{s} A_j Y_{t-j} + u_t$$

where the vector Y_t contains the contemporary values of the endogenous variables, each of which depends on its own lagged values and on contemporaneous and lagged values of the other variables. A_j are the coefficient matrices of the lagged values. The components of the residuals u_t are supposed to be uncorrelated, i.e. "clean" of those contemporaneous effects that are already included in the coefficient matrix A_0 ("orthogonalized residuals"). Multiplying both sides by A_0^{-1} yields the conventional VAR form

$$Y_t = \sum_{j=1}^{s} (A_0^{-1} A_j) Y_{t-j} + (A_0^{-1} u_t) \, with \, E(u_t u_\tau^{'}) = \begin{cases} I \; if \; t = \tau \\ 0 \, otherwise \end{cases}$$

that might be rewritten as

$$Y_t = \sum_{j=1}^{s} \Phi_j Y_{t-j} + \epsilon_t \, with \, E(\epsilon_t \epsilon_\tau^{'}) = \begin{cases} \Sigma \; if \; t = \tau \\ 0 \, otherwise \end{cases}$$

where consistent estimators of Σ and the Φ_j's are easily obtained by running OLS regressions equation by equation.

Additionally, estimation of A_0^{-1} is necessary to recover the contemporaneous response of the variables to orthogonalized shocks. However, as this requires estimation of an additional number of parameters, the system is not identified. A sufficient condition to reduce the amount of parameters is to restrict the VAR model by imposing lower triangularity of the matrix A_0^{-1} from using a Cholesky decomposition $\Sigma = A_0^{-1} A_0^{-1'}$. Multiplying the residuals by a lower triangular matrix implies that, given a particular ordering inside the vector Y_t, each variable is allowed to react within the current period to a shock in any of the variables of a higher ordering, while it is completely unresponsive to shocks in variables that are lower in the ordering. In this context, yearly demographic variables seem to fit the framework almost ideally as it can be clearly distinguished between contemporaneous and lagged effects. In the last chapters it was argued that childbirth rarely takes place in the same year as the fertility decision, in particular due to a pregnancy lag. Since this natural lag prevents it from being contemporaneously effected by death rate and GDP per capita, birth rate is the only plausible candidate to be the first variable in the vector Y_t. Furthermore, the death rate is placed as second variable to preserve the possibility of contemporaneous effects of GDP per capita growth on moratlity by an extended operation of the PoDR and the PoLD, which have so far been assumed to merely operate through the birth rate only. Finally, it is assumed that a change in GDP per capita does not affect the death rate in the same year, while a delayed negative effect retains the possibility of an endogenized mortality, yielding the following system for estimation:

$$\begin{pmatrix} b_t \\ d_t \\ g_{y_t} \end{pmatrix} = \Phi_1 \begin{pmatrix} b_{t-1} \\ d_{t-1} \\ g_{y_{t-1}} \end{pmatrix} + \Phi_2 \begin{pmatrix} b_{t-2} \\ d_{t-2} \\ g_{y_{t-2}} \end{pmatrix} + \Phi_3 \begin{pmatrix} b_{t-3} \\ d_{t-3} \\ g_{y_{t-3}} \end{pmatrix} + \begin{pmatrix} \beta_1 & 0 & 0 \\ \beta_2 & \beta_3 & 0 \\ \beta_4 & \beta_5 & \beta_6 \end{pmatrix} \begin{pmatrix} u_{b_t} \\ u_{d_t} \\ u_{g_{y_t}} \end{pmatrix} \tag{7.1}$$

7.1.6 Impulse response analysis

To find evidence for the classical growth model, the suggested linear relations should be approximately recovered by applying impulse response analysis to the above restricted VAR(3) model. To this end, nine orthogonalized impulse response functions are computed by shocking the error terms of each variable's equation by one standard deviation. The initial shock instantly affects the assigned contemporaneous variables and subsequently propagates through the system. Since childbirth is, as a response to shocks in death rate and GDP per capita growth, most arguably spread over a number of years, it is reasonable to expect *accumulated* orthogonalized impulse response functions to yield more pronounced and significant effects. On the other hand, as the period in question should not exceed the short term, a time horizon of more than five years seems inappropriate, granting that the fertility decision is usually made after four periods and that a longer horizon will not provide additional information. To test for significant effects, two hundred bootstrap replications are used to generate 95% confidence intervals. The size of the shocks will be given by the standard deviation of the corresponding variables.

If the considerations made in chapter three are correct, the causal relationships given by the estimated cumulative orthogonalized impulse response functions (coirfs) following a shock in the corresponding variable should be of the form

$$\begin{pmatrix} coirf_b \\ coirf_d \\ coirf_{g_y} \end{pmatrix} = \begin{pmatrix} high\,persistence\,(+) & short\,run\,(+)^1 & short\,run\,(+)^2 \\ (x) & high\,persistence\,(+) & (x) \\ contempor.(-)^3, long\,run\,(+)^4 & (x) & low\,persistence\,(+) \end{pmatrix} \begin{pmatrix} shock_b \\ shock_d \\ shock_{g_y} \end{pmatrix} \tag{7.2}$$

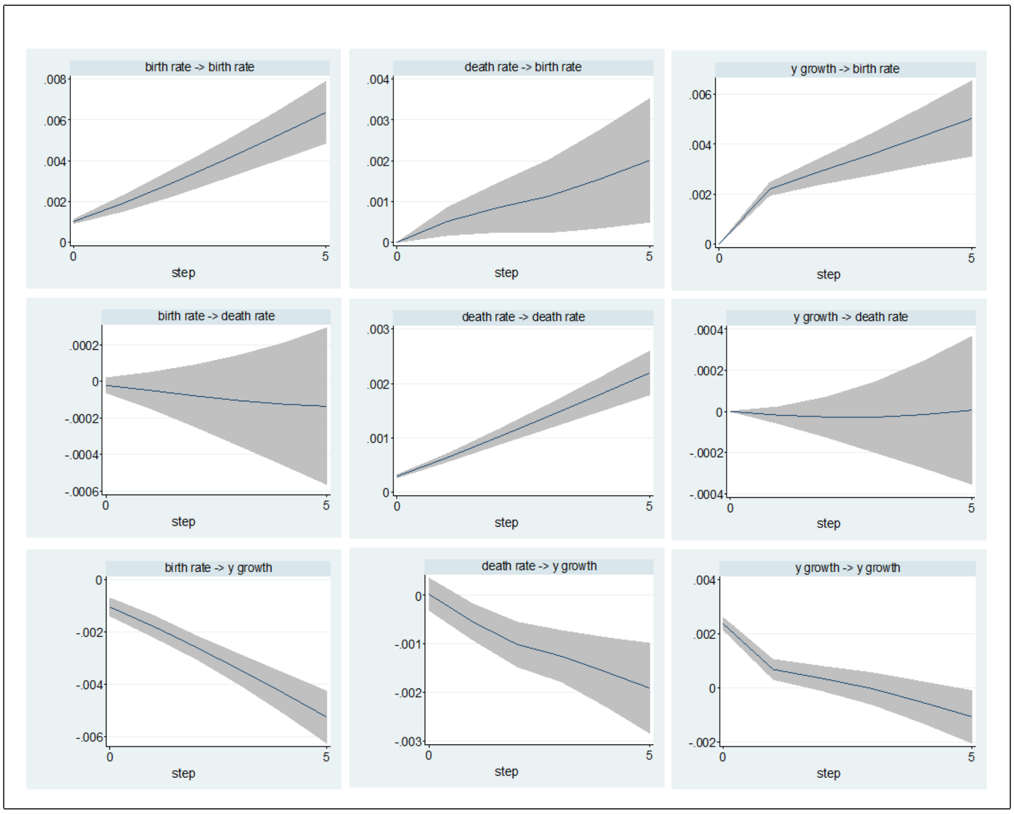
where $(+)^1$ is expected to display the positive average effect of the PoG and $(+)^2$ to capture the positive average effects of the PoP. $(-)^3$ is supposed to reflect the negative effect of the PoDR. This relation exists by definition and the effect will be observed as long as it is not outweighed by the impact of the PoLD. As was mentioned, $(+)^4$ will not be captured sufficiently well to account for the positive effect of the PoLD and its presence even poses a threat to a clear identification of the PoDR. Consequently, since PoDR and PoLD may not be clearly identifiable within the VAR(3) framework, their existence can certainly not be falsified in this framework. Therefore, the subsequent investigation will focus on the evaluation of the hypotheses of the PoG and the PoP, the classical theory of population. The existence of the PoLD and the PoDR will be conjointly estimated in the next chapter by applying on OLS estimation to equation 3.6 Persistence effects are expected to be measured for the variables birth rate and death rate, much less for GDP per capita growth. The remaining three impulse response functions denoted (x) will also be estimated to capture further potential effects by which the classical model might be extended ex post. The resulting impulse responses should be interpreted with care, as the effects of the PoG and the PoP are supposed to be time–varying, whereas the estimation can merely give average results over the whole period in question.

7.2 Estimation Results

7.2.1 Estimation of the simulation

As a very useful reference point, it is advisable to first run a VAR(3) estimation on the above simulation given by Figure 6.2, i.e. on the system 6.1 using the corresponding calibration.[7] The coirfs resulting from this estimation are expected to deliver a benchmark against which the ensuing real samples might be compared. The universal average effect of the PoG seems well exposed in the upper central graph of Figure 7.2.

Figure 7.2: Simulating the classical growth model: Coirf matrix based on a VAR(3) model.



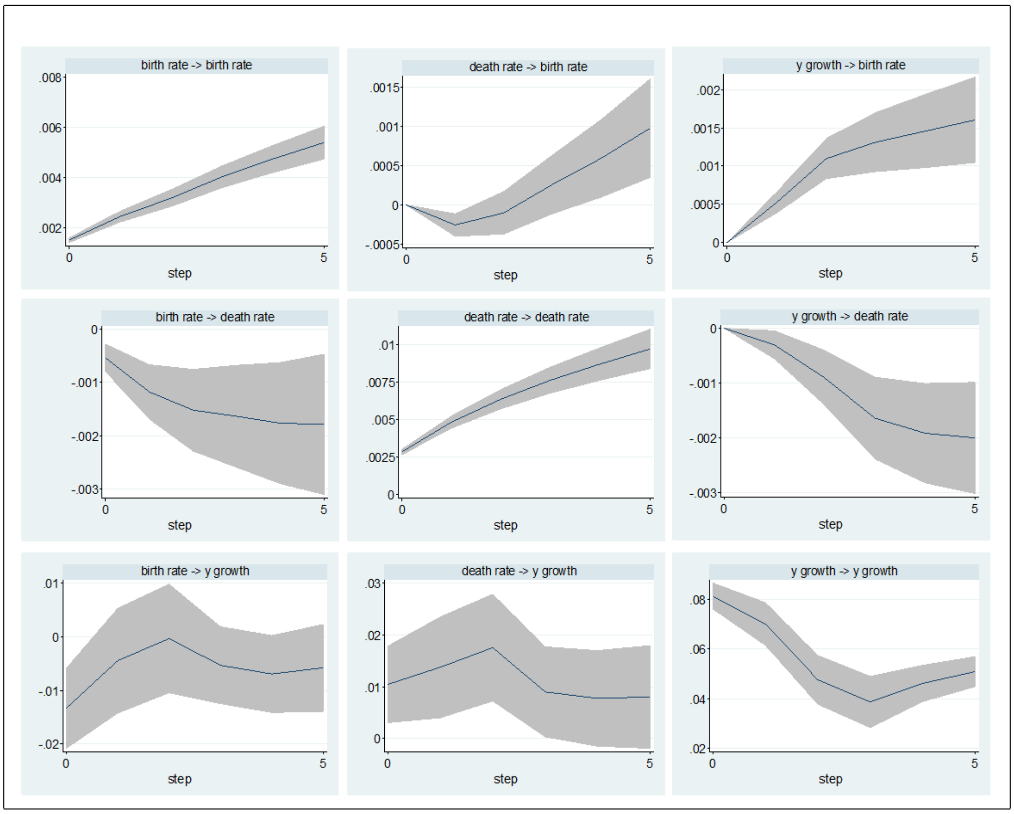
Likewise, the average effect of the PoP appears quite nicely depicted in the upper right graph. Both effects are statistically significant, with the PoG on a five percent level and the PoP on a one percent level. As suggested, the positive lagged effect of the PoLD of birth rate on GDP per capita growth after one generation cannot be captured in the bottom left graph, when using a maximum lag length of three. Instead, the contemporaneus annual effect of the PoDR is significantly displayed in the same graph at period zero, since it is not distorted in this simulation by a permanently operating effect of the PoLD.

[7] To provide some additional variation on the variable death rate, the right hand side of the second difference equation of eq. (4) is supplemented by adding an annual shock $\epsilon_t \sim U(-0.005, 0.005)$.

7.2.2 Britain 1541–2010

The computed coirfs from running the VAR(3) model on the British data are displayed in Figure 7.3. The effect stemming from the PoP is roughly in line with that of the simulated model. The reaction of birth rate to a shock in GDP per capita growth is already positive and significant on a one percent level in the first period, indicating a quick fertility adjustment, and accumulates in magnitude over the subsequent periods. With regard to the PoG, a death rate shock does not induce birth rate to react after one period. After four years, however, the positive effect becomes statistically significant on a five percent level, providing evidence of a positive causal relationship and pointing to a lagged fertility decision.[8]

Figure 7.3: Britain 1541–2010: Coirf matrix based on a VAR(3) model.



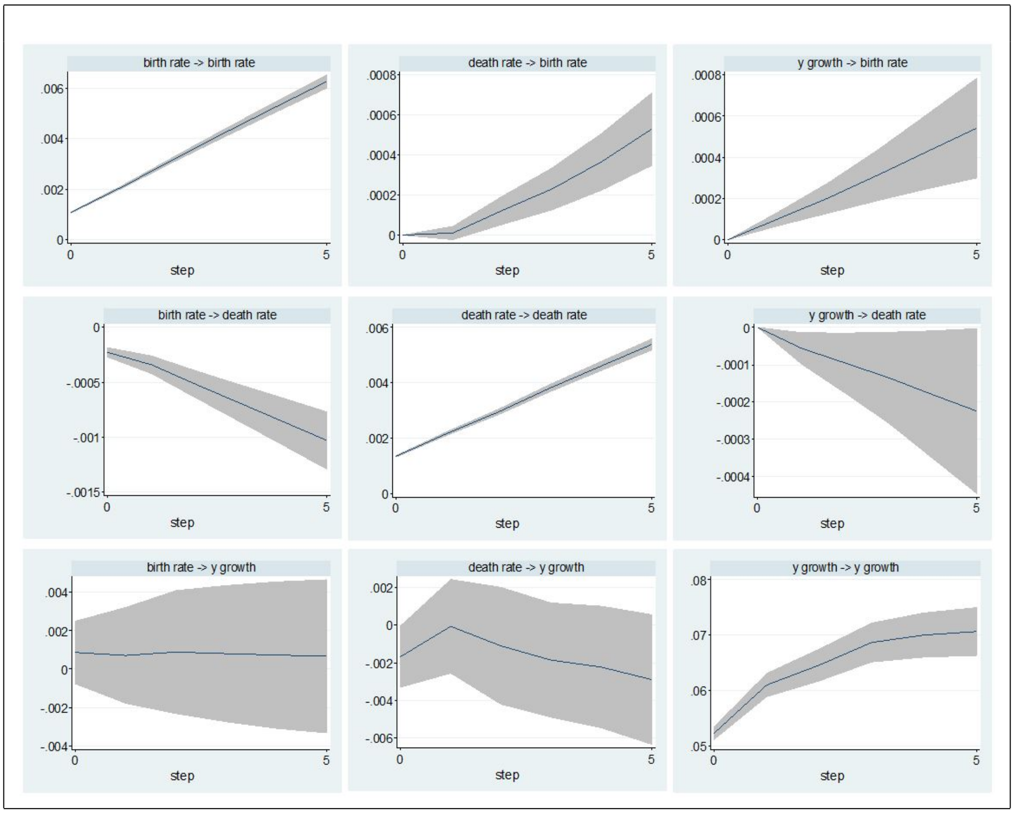
Although the effect of the PoDR is significantly measured in year zero after a shock in birth rate, the subsequently displayed response until year five might have been distorted by the PoLD. While in the simulated impulse response the negative contemporaneous effect of the PoDR seems to have accumulated over the subsequent periods, this accumulation is neutralized in the British sample, suggesting that a positive (potentially long-run) effect of birth rate on GDP per capita growth has been captured by additional residual correlation.

[8] In the simulation of chapter three, a lag of one year was used. However, this assumption can be easily relaxed by assuming a fertility lag of up to four years.

7.2.3 Stacked model using international data

Mitchell's international historical statistics offer data on vital rates and GDP per capita for 94 economies. Out of these, 55 economies exhibit simultaneous data on the three variables over at least three consecutive years. However, with an average number of approximately 70 observations, the impulse responses of most of these 55 eligible countries can, when individually tested, not be expected to give sufficiently reliable evidence of the classical model. On the other hand, if they could be estimated collectively, the number of observations would rise to 3,911. For that purpose, the individual country–level data are stacked into one sample, leaving space for three "missing values" between sub–samples such that the last observation of the preceding and the first observation of the succeeding country are not related to each other.

Figure 7.4: 55 economies 1815–2010: Coirf matrix based on a VAR(3) model.



The resulting coirfs displayed in Figure 7.4 match those of Figure 7.3 relatively well. However, as the stacked sample includes observations between the years 1815 and 2010 only, while the English sample ranges from the year 1541 to 2010, the former is suspected to mainly include information on the regime of economic developement. It is therefore not surprising that the effect of the PoG is remarkably pronounced compared to the British model, while that of the PoP is smaller in magnitude, suggesting time-varying effects. Moreover, the higher persistence in the variable GDP per capita growth rate is in line with the suggestion that its mean was close to zero during the regime of stagnation and increased sustainably during the regime of development. However, as in the British sample, it remains puzzling why the fertility decision seems to be lagged by an additional year in the case of the PoG as compared to the effect of the PoP. Finally, in contrast to the simulation where we have merely employed the effect of one cohort on the subsequent GDP per capita growth, the effect of the PoDR seems to have been neutralized by the PoLD such that no significant effects can be observed. In any case, both analyses of the British historical data as well as of the more recent international data record qualitative and quantitative evidence of a universal pattern of birth rates, death rates and GDP per capita growth and strong support for the universality of the PoG and the PoP.

7.3 Robustness Checks

7.3.1 Alternative British dataset

In the following, robustness checks are conducted with respect to alternative data, logged variables, alternative orderings of the variables, country–specific estimations and the hypothesis of time–varying coefficients. Firstly, since Clark's macroeconomic aggregates for Britain have not been fully accepted by many economic historians, the VAR(3) model will also be applied to the Broadberry et al. dataset of GDP per capita. The resulting coirf–matrix displayed in Appendix 11.3 yields a very similar pattern to the matrix using the Clark dataset. Likewise, the effects of the PoG and the PoP are significant after four periods, while the immediate effect of the PoDR and the lagged effect of the PoLD cannot be identified. Although, therefore, the Broadberry et al. data suggest no stagnation during the seventeenth and eighteenth century, they confirm the observability of the classical (Malthusian) theory of population as well.

7.3.2 Using logged birth rates

Furthermore, in the theory of production it has been argued that there exists an exponential relationship – instead of the linear relationship employed in our simulations – between birth rate and GDP per capita growth. This feature can be accounted for by replacing the absolute value of the birth rate with the logged birth rate and estimate the VAR(3) model on the British data as well as on the international stacked data again, based on this new variable. As is shown in Appendix 11.3, this re–estimation does not greatly change the above results, as the PoG and PoP remain well observable even when using logged birth rates.

7.3.3 Using an alternative variable ordering

Thirdly, we evaluate the impact of a different variable ordering on the PoG and the PoP. The resulting coirfs are again displayed in Appendix 11.3.

Firstly, if the ordering of the three variables is so arranged that the birth rate still precedes the death rate ($\begin{pmatrix} b \\ g_y \\ d \end{pmatrix}$ *or* $\begin{pmatrix} g_y \\ b \\ d \end{pmatrix}$), the coirfs from the corresponding autoregressions remain relatively unchanged. Accordingly, it does not play a large role, whether GDP per capita growth is allowed to affect the birth rate in the same year, since the correlation (the contemporaneous effect) between the two variables is in any case very small. However, when making use of the possibility that the death rate is allowed to affect the birth rate contemporaneously, as is shown in the graphs resulting from the different orderings ($\begin{pmatrix} g_y \\ d \\ b \end{pmatrix}$, $\begin{pmatrix} d \\ b \\ g_y \end{pmatrix}$ *and* $\begin{pmatrix} d \\ g_y \\ b \end{pmatrix}$), the effect of the PoG can no longer be identified. Although the effect of the PoP is well in place, a strong negative (contemporaneous) correlation between birth rate and death rate outweighs the PoG. Consequently, we argue that the original ordering with the birth rate as first variable and the death rate as second variable is valid and crucial in measuring the PoG.

7.3.4 Estimation on a Country-level

Fourthly, the robustness of the PoG and the PoP will be checked on a country–level. Since the number of parameters of the above VAR(3) model amounts to 36 when including a vector of intercepts, it does not seem reasonable to evaluate economies with less than seventy observations on death rate, birth rate and GDP per capita growth rate. For this reason, merely all available country–level time series providing at least seventy observations on the three variables are employed for empirical evaluation. The remaining twenty countries, including Britain, are listed in Appendix 11.3. Following the above estimation procedure, it should be kept in mind that the resulting coirfs are naturally suspected to be less significant due to the smaller sample sizes. The correspondingly estimated coefficients of the PoG and the PoP after four years are displayed in Table 7.2, also including the two above British cases for comparison.

Astonishingly, in spite of the small sample sizes and with the exception of Denmark, each of the 19 economies display a positive coefficient in the case of the PoG as well as for the effect of the PoP in addition to the significant findings in the British cases. What is more, 17 out of 40 tests (excluding the British data by Broadberry et al.) display significant coefficients. These results suggest that even very small samples may be capable of providing evidence for the two principles sustaining our conclusions from the stacked model. For a more detailed examination of the stability of the effects, the national coirfs with five–year time horizon are also displayed in Appendix 11.3.

7.3.5 Time-varying effects

Finally, when accounting for the time-varying effects of the PoG and the PoP, the classical theory of population suggests that the PoG grows stronger whereas the PoP grows weaker during the transition to growth. Again following Nicolini (2007), a straightforward way to measure the evolution of the corresponding effects on the birth rate, which are supposed to increasingly respond to the declining level of the death rate, is to split up the British sample into an early period of economic stagnation and high mortality and a late period of economic growth and low mortality and to compare the respective coirfs. As, in accordance with the stylized facts, the growth take-off corresponds to the fertility decline, 1815 is chosen as the cutoff year, as it exhibits the maximum value and a structural break for birth rates. However, with the first sample

Table 7.2: Cumulative orthogonalized impulse response of b in % 4 periods after a one standard deviation shock in d and g_y. All 20 countries with more than 70 observations are displayed.

country	Principle of Population $g_y \rightarrow b(4)$	Great preventive Check $d \rightarrow b(4)$	obs
arg	0.039	0.018	97
au	0.018	0.077**	75
aus	0.035	0.115***	133
can	0.052	0.039	100
chil	0.097	0.064	99
den	-0.027	0.075***	179
fin	0.051	0.172***	147
fra	0.076***	0.059***	182
ger	0.148***	0.089***	122
hun	0.161***	0.038	72
ita	0.057	0.035	102
jap	0.005	0.066*	109
net	0.015	0.067**	91
nor	0.008	0.077***	128
nz	0.013	0.024	76
rom	0.034	0.043	79
spa	0.074**	0.003	91
swe	0.053	0.108***	146
swi	0.017	0.047**	78
ukc	0.146***	0.060**	467
ukb	0.092***	0.060***	467

*** indicates significance at 1% level, ** at 5% level, * at 10% level.

employing 271 observations and the second sample using 192 observations only, the outcome can again merely be considered as indicative evidence.

Figure 7.5: Britain: The evolution of PoG, PoP, PoDR and "positive checks". Upper sample 1541–1815. Lower sample 1815–2010.

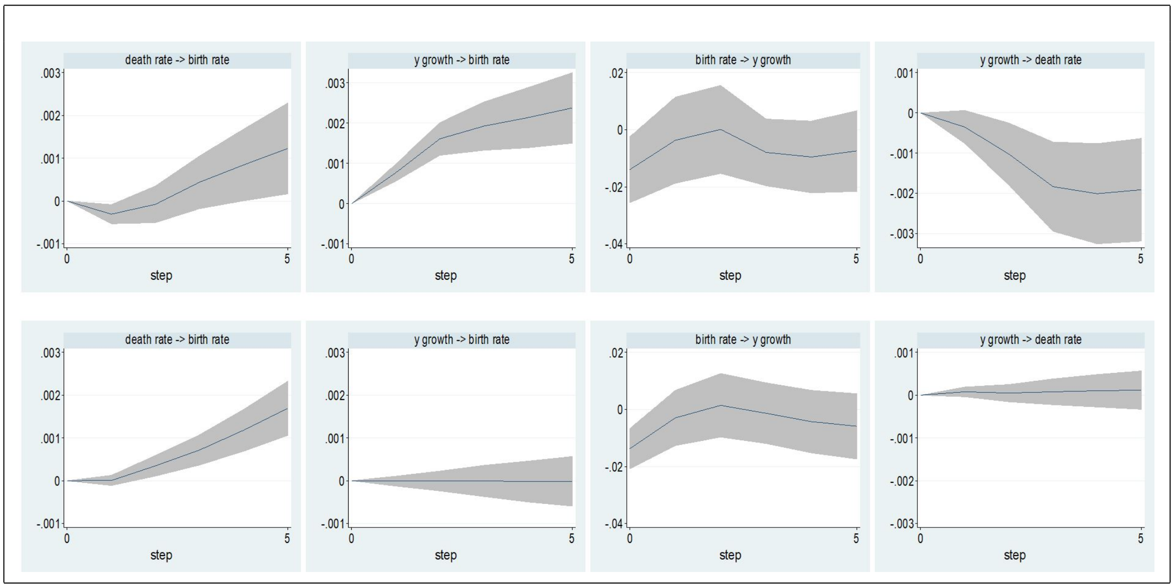


Figure 7.5 illustrates the evolution of the PoG and the PoP respectively from left to right. In the upper row, coirfs are given for the timespan 1541–1815, the bottom row displaying responses for the period 1815–2010. First, the effect attributed to the PoG grows in size and significance over time, suggesting an increasing direct mortality effect as is predicted from the theory. Secondly, the impact of the PoP sharply decreases over time. While indicating a strongly positive conversion rate of GDP per capita growth into birth rate in the early sample, the effect seems to entirely disappear in the later period. This diminishing income effect is in line with the increasing operation of indirect mortality effects. However, as it seems implausible to argue that the effect of the PoP completely vanished after 1815, the extremely low values of the coirf might be owed to distortions resulting from the small sample size. Thirdly, the relatively unchanged coirf that is supposed to capture the constant effects of the PoDR and the PoLD substantiates the theoretical predictions and sustains the robustness of the estimation method.

7.4 A Critical Note on the Prevailing Measurement of Preventive Checks in Cliometrics

Lastly, having found evidence of the existence of the short–run mechanisms suggested by classical growth theory, an additional coirf will be briefly interpreted, as it is regularly used in the prevailing empirical literature. The effect that might be important to consider arises from the statistically significant negative lagged response of death rate to changes in GDP per capita growth as is illustrated by the respective central right coirfs of Figures 7.3 and 7.4. So far, exogeneity of the death rate has been assumed to trigger the epidemiological transition. As a short–term relation, however, GDP per capita growth seems to have affected mortality even before the epidemiological transition, since this effect can be found to prevail in the early English data sample (see Figure 7.5, upper right graph) and to wear off in the late sample (see Figure 7.5, lower right graph).

Some authors (Nicolini 2007; Crafts and Mills 2009; Pfister and Fertig 2010; Fernihough 2012; Herzer et al. 2012; Moller and Sharp 2014; Rathke and Sarferaz 2014; Edvinsson 2017) have argued in favor of complementing "Malthusian effects" in the sense that higher productivity not only enhances fertility, but at the same time operates towards lower mortality. These authors regard the statistically significant positive effect of GDP per capita growth on birth rate – corresponding to the effect of the PoP in this work – as evidence of "preventive checks" in general, which are not to be confused with the PoG. Accordingly, if income was observed to raise births on average, it is a sign that reproduction has formerly been suppressed by preventive fertility behavior. Equally, they hold the apparent negative causal relationship between GDP per capita and death rate to universally reflect "positive checks". Their idea is that whenever living standards would fall below a subsistence level, the positive checks are supposed to increase the death rate as a general result of individuals heavily competing for the remaining resources.

This effect deserves attention and could be added to the simulation to complement the mechanism of stagnation by providing another channel of population growth. In this paper, the modeling of the effect of conventional "positive checks" has been disregarded for two reasons. Firstly, it does not provide explanatory power for the mechanism of growth, since the positive checks are thought to disappear at the same time as GDP per capita rose above subsistence level. When modeling and evaluating the growth regime, it is regarded to be sufficient to focus on the steady decline of fertility as the crucial factor contributing to the population slowdown inducing the breakout from stagnation. Secondly and more importantly, the current conventional cliometric interpretation of preventive and positive checks is at odds with Malthus' definition that

> the preventive check is perhaps best measured by the smallness of the proportion of yearly births to the whole population,[9]

i.e. by the level of the birth rate and that

> the positive checks to population [. . .] include every cause [. . .] which in any degree contributes to shorten the natural duration of life,[10]

which are best measured by the level of the death rate. Consequently, the preventive checks ought not to be measured by the causal relationship running from GDP per capita to fertility, which is reserved for the PoP. Instead, it might be very generally concluded that a low birth rate is a sign of the operation of preventive checks, whereas a high death rate reveals the operation of positive checks. Naturally, this implies an important Malthusian insight that has already been hinted at — that the regime of stagnation is characterized by high mortality and the regime of development by low fertility.

[9] Malthus (1826), book II, chapter XI.
[10] Malthus (1826), book I, chapter II.

7.5 Concluding Remarks on the Empirics of the Theory of Population

Since it was considered insufficient to construct a model that merely fits the stylized historical facts of stagnation and growth, the operation of the classical principles has in this chapter been evaluated collectively to avoid the reasonable impression of "reverse engineering" in our simulations. To this end, a simple VAR estimation provided a way to establish evidence of the suggested classical short–run relationships by employing orthogonalized cumulative impulse response functions derived from two historical samples, based on approximately 4,500 observations of annual national data on birth rate, death rate and GDP per capita growth. In those cases, in which the principles were a priori supposed to be measurable, in particular for the PoG and the PoP, the impulse responses yielded strong support for the theory. Additional robustness tests conducted with regard to country–specific effects and time–varying coefficients were generally in line with the classical principles. Also, it has been suggested that recent publications might require reconsideration regarding the use of "positive checks" and "preventive checks", as they seem to be at odds with Malthus' original terminology. Notwithstanding the empirical validation of the PoG and the PoP, an important shortcoming of this work lies in the omission of the effects to be observed from the principle of diminishing returns and the principle of labor division. Until they can be measured, the classical unified growth model cannot be said to have been fully confirmed by the data. We will therefore now turn closer to the first equation of system 6.1. For future research, the most obvious extension of the classical unified growth model that has already been hinted at, would be to endogenize mortality. Here, it might be advisable to model a linear effect of a logged variable GDP per capita growth on death rate to account for an increasing difficulty to generate a higher life expectancy.

Chapter 8

Measuring the Neoclassical Theory of Production

In section 7.1 it was stated that a VAR(3) model is not able to capture the effects from the PoLD, since this principle can only be observed over a longer time frame than the suggested four years. In our simulations we have assumed that a larger birth rate will induce the corresponding cohort to increase GDP per capita *only at one point in time* – exactly fifteen years after they have entered the labor market. However, when trying to measure the operation of the PoLD in reality, we have to take the following two problems into account. Firstly, every individual tends to generate increasing per capita growth not only during the first year of his working life, but over his complete lifetime due to the accumulation of human and physical capital. This means that the positive effect of population growth on GDP per capita is spread over the lifespan of the newborn individuals, which may even require the use of up to one hundred lags for empirical investigations. Secondly, the PoDR operates at the same time continuously to mitigate the increase in GDP per capita. Consequently, the effects of these two principles – the positive and the negative effects of population growth on GDP per capita growth – are constantly intermingled in all our employed data. While we have seen that diminishing returns were significantly measured by the corresponding coirf of the simulation (left bottom graph of Figure 7.2), this effect could not be found in the stacked data sample (see left bottom graphs of Figure 7.4). Due to the inability of the VAR–approach to estimate long–run relationships (which would require an enormous sample size to account for an enormous amount of parameters), we will in the following employ a simple growth regression approach to evaluate the neoclassical theory of production by measuring the PoLD and PoDR conjointly.

8.1 Methodology

To support the advanced theory of production, the subsequent empirical exercise will focus on the estimation of steady–state equation 3.6, as stated in chapter three,

$$\frac{y_t^*}{y_{t-j}^*} = \left(\frac{b_{t-j}}{b_t}\right)^{\frac{\gamma}{1-\gamma}},$$

and proceed as follows. To account for an appropriate $j-$value, we will in the following theoretically justify a presumed time span between steady states. To find evidence of a causal relationship of birth rate on GDP per capita growth, we will recover the remaining parameter γ by estimating the expression $\frac{\gamma}{1-\gamma}$. Based on the assumption of constant returns to scale and indicating the supposed negative long–run effect of a change in the birth rate on productivity, the estimated value of the parameter γ is expected to lie within the range $(0, 1)$. As a reference point, Cobb and Douglas (1928) estimated the production elasticity of capital of the (original) labor model to be $\alpha \approx 0.25$. More recently, Mankiw et al.'s (1992) estimations suggested a parameter value $\alpha \approx 0.33$ and most conventional calibrations assume a production elasticity of capital within the range $(0.25, 0.33)$. Nonetheless, when withdrawing the variable human capital from the variable labor and adding it to the variable physical capital while treating the remaining "unskilled labor" as population, Mankiw et al.'s estimator rises to $\alpha \equiv \gamma \approx 0.66$. Such an estimator would suggest a much higher exponent $\frac{\gamma}{1-\gamma}$ in equation 3.6 and correspondingly a larger leverage effect of changes in the birth rate on GDP per capita. In fact, if the following estimation confirms the conjecture that $\gamma \approx 0.66$, population growth must be considered to have a much greater impact on economic development than is usually suspected.

Ideally, once we have estimated a consistent parameter value, we can confirm or reject the time frame presumption for j. To regress equation 3.6, we will employ the usual ordinary bivariate least squares (OLS) method. Due to the fact that the OLS estimator is the best *linear* unbiased estimator (BLUE), equation 3.6 will be linearized by taking logs of both ratios, yielding the approximate growth of GDP per capita as explained variable and the inverted approximate growth of birth rate as explanatory variable (see equation 8.1).[1] OLS estimation is in this case a valid approach, since the variable birth rate is viewed as the (independent) source of all value in GDP per capita. This also implies that the additional use of an intercept or fixed effects to account for unobserved effects is not necessary.

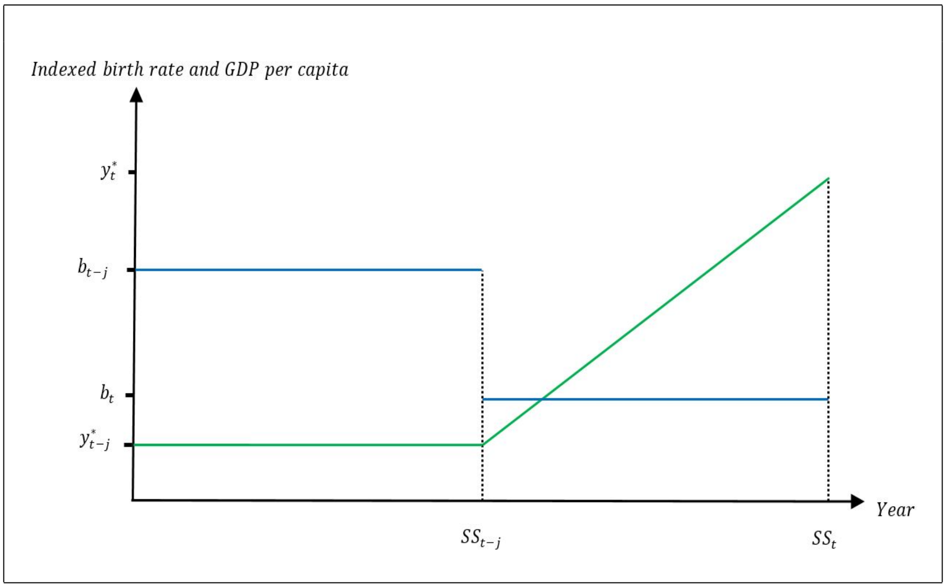
$$ln\left(\frac{y_t^*}{y_{t-j}^*}\right) = \beta\ ln\left(\frac{b_t}{b_{t-j}}\right)\ with\ \beta = -\frac{\gamma}{1-\gamma} < 0. \tag{8.1}$$

However, when regressing equation 8.1 it must be noted that an ideal determination of β would only be realized under a comparison of the two steady states $y_t^*$and y_{t-j}^*. Since the transition from one steady state to another might take a certain number of years after a shock in the variable birth rate has occured, we have to account for this transitional period by using an appropriate value for j. Although the birth rate is expected to affect productivity immediately negatively

[1] For an alternative methodology including a measurement of convergence see Mankiw et al. (1992).

through b_t, its positive effect of labor division is realized with a lifetime–delay through b_{t-j}. To account for the latter effect, it has been stated that a newborn cohort will raise productivity only after it has generated the ability of unskilled labor with a maturity lag of one "generation" of ϕ years, and it seems plausible to continue to assume a period of at least15 years. In addition, broad capital accumulation by way of "dexterity" and the "invention of machines" will be assumed to take place over the whole working life of a cohort, i.e. we add a maximum amount of $\psi \approx 50$ years.[2] Thus, since the above combined gains are probably fully achieved after $j = \phi + \psi$ years, we assume a maximum accumulation period of $j \approx 65$ years to account for the transition between steady states. Consequently, ideal results from an OLS estimation can only be expected if we could employ time series of GDP per capita and birth rate over a time horizon of $2j \approx 130$ years where the birth rate stays constant for the first j years, changing abruptly to another level (treatment) and remaining constant on the new level for another j years, as is exemplified in Figure 8.1. After the birth rate has changed, GDP per capita is predicted to react positively over the latter period (treatment effect).

Figure 8.1: Illustration of two time series required for an ideal estimation of equation (11).



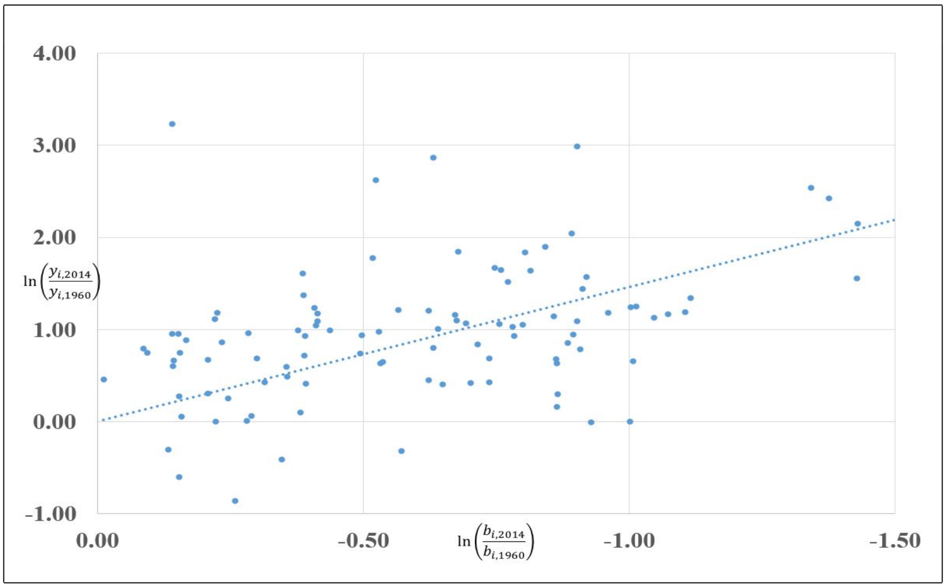
[2] For a more extensive lag model, see e.g. Becker and Murphy (1992) or Liso et al. (2001).

8.2 Estimation of γ and j

8.2.1 Estimation of γ

Firstly, to get an idea about the global empirical relationship between birth rate and GDP per capita, available aggregate data series provided by the World Bank are displayed in Appendix 11.4. Here, we observe a relatively steady decline of the birth rate as well as a parallel rise in GDP per capita over the period 1960–2014. The corresponding calculation of the aggregate parameter yields $\beta^g = -2.0$ (see column (1) of Table 8.1) and we conclude that an average birth rate reduction by 1% is connected with an average rise of GDP per capita by about 2%. Secondly, using country-level data, we find that the birth rate decreased during the period 1960–2014 in each of the 104 available series (see Appendix 11.4 for a list of countries studied).

Figure 8.2: Scatterplot of 104 countries comparing (negative) growth of birth rates (x–axis) and growth of GDP per capita (y–axis) between 1960 and 2014.



Source: World Bank (2018).

Consequently, we would tend to expect a rise in GDP per capita in every country at least until the year 2014, which we will thus use as reference year to represent the terminal steady state y_t^*. To find evidence for the expected relationship, we first plot the dependent variable (productivity growth) against the independent variable (inverted growth of birth rate) over the complete 55–year period for all 104 countries with available data (see Figure 8.2) and then estimate the OLS coefficient of equation 8.2.

$$ln\left(\frac{y_{i,2014}}{y_{i,1960}}\right) = \beta\ ln\left(\frac{b_{i,2014}}{b_{i,1960}}\right)\ \ for\ i = country\ 1,\ ..,\ country\ 104 \tag{8.2}$$

The results are displayed in column (2) of Table 8.1. The R–squared of 0.698 indicates that the greater part of the variation in GDP per capita is explained by the variation in the birth rate.

The absolute magnitude of the coefficient is somewhat smaller than the aggregate coefficient in column (1), which is certainly at least partly due to the fact that the OLS approach does not weigh countries according to their population size.[3] Nonetheless, since the coefficient is highly significant, we have found some evidence of the true parameter β lying approximately within the range $[-2.02, -1.56]$ over the observed time span 1960–2014. However, since the above theory maintains that a change of the birth rate affects GDP per capita over the following $j \approx 65$ years, a calculation employing a time horizon of merely 55 years may underestimate the magnitude of the aggregate coefficient, suggesting a somewhat higher true value.

Consequently, in order to extend the maximum time span for j toward 65, we turn again to the Mitchell (2013) database and estimate the corresponding coefficient for a sample including all 34 countries providing data on GDP per capita and birth rate for the year 1949 as well as the year 2014 (column (3) of Table 8.1). The greater magnitude of this "long-run" coefficient as well as the higher R-squared seem to confirm our expectation with regard to the time horizon.

Table 8.1: Calculated and estimated coefficients.

	$ln\left(\frac{b_{i,2014}}{b_{i,t}}\right)$ (1)	$ln\left(\frac{b_{i,2014}}{b_{i,t}}\right)$ (2)	$ln\left(\frac{b_{i,2014}}{b_{i,t}}\right)$ (3)	$ln\left(\frac{b_{i,2014}}{b_{i,t}}\right)$ (4)
$ln\left(\frac{y_{i,2014}}{y_{i,t}}\right)$	-2.02	-1.56***	-2.19***	-2.16***
		(.100)	(.134)	(0.178)
R^2		0.70	0.79	0.94
t	1960	1960	1949	1901
$\# i$	1^a	104^a	34^b	10^b

*** indicates significance at 1% level. Standard errors are reported in parentheses.
Sources: a=World Bank aggr. global data (2018), b=Mitchell (2013)

8.2.2 Estimation of j

To show that the true parameter j actually centers around 65 years, we will employ the World Bank as well as the Mitchell data series conjointly and display the evolution of the coefficient β for increasing j. As is shown in Figure 8.3, the coefficient remains significant for all j–values. As expected, by increasing the transitional time span j, the coefficient tends to increase as well until settling at a value of approximately –2.0 after 60–65 years. Thereafter, the coefficient remains relatively constant at an average value of –2.0 and the 95% confidence interval roughly within the boundaries $[-3.0, -1.5]$ for $j > 60$.[4]

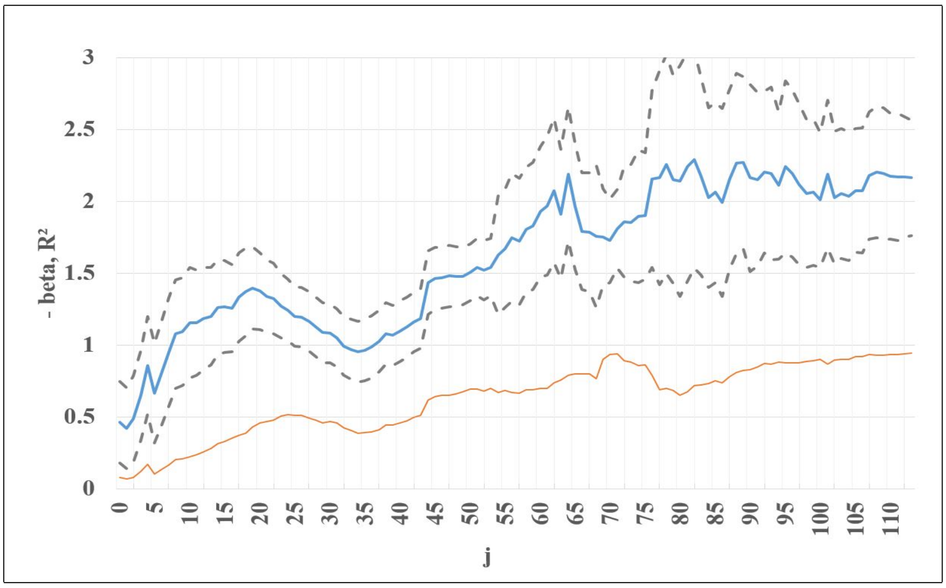
On the one hand, these observations confirm the predicted strong impact of birth rate changes on GDP per capita. Since the displayed R–squared tends to steadily increase over time and

[3] If, for example, China and India were assigned a weight according to their population size, the absolute value of the coefficient would be larger.

[4] Constancy is here defined as a linear trend with slope parameter$< |0.002|$ over the corresponding time span.

displays a value of 0.94 for the longest period $j = 112$ (column (4) of Table 8.1), it appears that changes in the birth rate are capable of explaining over 90% of the subsequent GDP per capita growth. On the other hand, they provide evidence of the idea that the full effect of changes in the birth rate is achieved after approximately 65 years.

Figure 8.3: Magnitude of OLS-estimatorβ (blue) with 95% confidence intervals (gray) and R^2 (yellow) with increasing time span j.



As a result, using the average coefficient value for $j > 60$ as a benchmark, the theory suggests that a 1% decline in the birth rate causes on average a 2% increase in GDP per capita *growth* over the subsequent 65 years. These values imply that a birth rate reduction from 4% to 1% tends to raise production per capita by the factor sixteen.[5] Finally, the production elasticity of broad capital $\gamma = \frac{\beta}{1+\beta}$ can be calculated to lie approximately within the 95% confidence interval $[0.6, 0.7]$, confirming Mankiw et al.'s (1992) estimation results.

8.3 Robustness on a Country–Level

To indicate the robustness of the above approach, we may report the country–specific results, as has been done in the case of the theory of population. In Table 8.2, the national parameters γ_i are displayed for the 34 states exhibiting data for 1949 and 2014. As expected, all parameters display positive values between zero and one. Except for four Latin American economies, the value of the calculated parameters lies between 0.50 and 0.86. This range of variation seems reasonable given that our population approach does not control for the international division of labor and the related foreign capital movements. The average national parameter $\bar{\gamma}_i$ of exactly 0.66 might then be interpreted to account for this mobility of capital by suggesting that withdrawing capital from one country might speed up the growth process in another one. For example, if the effect of

[5] The possibility of reverse causality is dealt with in section 8.4.

Table 8.2: Calculated parameter γ for the 34 economies exhibiting data for 1949 and 2014 measuring the combined effect of the PoLD and the PoDR

country	γ_i	country	γ_i	country	γ_i
arg	0.38	gre	0.72	pan	0.80
aus	0.80	hon	0.30	phi	0.81
au	0.84	ind	0.80	por	0.59
bel	0.76	ire	0.86	spa	0.66
can	0.64	isr	0.88	sri	0.68
chil	0.53	ita	0.66	swe	0.75
col	0.55	jam	0.70	swi	0.62
cyp	0.82	jap	0.62	uk	0.75
den	0.70	mex	0.53	uru	0.70
egy	0.84	net	0.69	ven	0.22
els	0.35	nor	0.75		
fin	0.63	nz	0.50		

birth rate on GDP per capita growth is found to be larger than 0.80, the growth process might have been positively affected by additional foreign capital inflows, speeding up the process of "inventing machines" suggested by the PoLD, whereas the outflow of capital in Latin American economies might have slowed down their rate of capital accumulation usually resulting from the division of labor.

8.4 A Popular Objection: Reverse Causality?

As has already been discussed in section 4.4, the observed correlation between birth rate and GDP per capita has prompted large academic circles to believe that rising productivity generally induces individuals to lower their fertility, since running a regression for the (inverted) equation

$$ln\left(\frac{b_t}{b_{t-j}}\right) = \alpha\ ln\left(\frac{y_t}{y_{t-j}}\right) \tag{8.3}$$

would naturally yield an inverted significant coefficient $\alpha = \frac{1}{\beta}$. As it is a quite popular objection, the inevitable reply will, even in this work, be that the observed negative effect of fertility on productivity is merely another illustration of a statistical correlation being misread as a causal effect.[6] However, this hypothesis must be empirically rejected as well for the following reason. Firstly, due to a fertility decision lag and a pregnancy lag, a contemporaneous effect of y_t on b_t can barely exist. Eventually, since we stated that birth rate has a delayed effect on GDP per capita, we may also test for a delayed effect (l) of GDP per capita on birth rate of the form

6 See, for example, Becker (1981).

$$ln\left(\frac{b_t}{b_{t-j}}\right) = \alpha\ ln\left(\frac{y_{t-l}}{y_{t-l-j}}\right). \tag{8.4}$$

However, although GDP per capita has steadily increased over the 20th century in developed economies, we observe – beginning in the 1970s – a constant birth rate in those countries.[7] If GDP per capita would indeed had a negative impact on birth rate, we should instead observe a further declining birth rate after 1970, which is obviously not the case. Further reading on the empirical rejection of the Becker–Hypothesis is provided by Galor (2011).

8.5 Concluding Remarks on the Empirics of the Theory of Production

The above findings strongly support the validity of the employed Solow–Model without "technology" and thus yield strong evidence of the simultaneous operation of the classical principles of labor division and of diminishing returns provided by equation 3.6. If there exists an equally strong negative contemporaneous effect (PoDR) and a positive lagged effect (PoLD) of population growth on GDP per capita growth over the subsequent periods, a change in the population growth rate will obviously negatively effect GDP per capita growth, as has been impressively confirmed by the data as follows.

Evaluating annual data on 104 countries over a period of 55 years, 34 countries over a period of 65 years and 10 countries over a period of 114 years, our estimations imply that a birth rate reduction from 4% to 1% – as is often observed in developed countries – raises production per capita by the factor sixteen. If these results are correct, the historically observed decline in fertility can account for the largest part of the historically experienced sustainable rise in production per capita. Moreover, our estimations suggest that the production elasticity of broad capital lies in the range (0.60, 0.75), suggesting a production elasticity of population in the range (0.25, 0.40), which has often erroneously been calculated to lie in the interval (0.66, 0.75). Although employing a quite different approach, the results are roughly in line with those of Mankiw et al. (1992) and Ashraf et al. (2013), providing supportive evidence of the neoclassical growth model. Further research will be required to confirm the idea that an appropriate form of the aggregate production function is approximated by $Y(K,H,N) = K^{1/3}H^{1/3}N^{1/3}$.

A calculation of γ may alternatively be conducted by using the production exhaustion theorem to determine the income share of population. To this end, we would have to employ average unskilled labor wages or minimum wages with regard to the whole population (which may in fact be termed "geographical wages" or "population wages") and to compute their income share (1- γ) on total GDP. This share should be found to lie in the interval (0.25, 0.40). Moreover, while this empirical exercise provided a relatively simple approach to productivity, adhering to Cobb and Douglas' "method of attack," the model might be extended by accounting for a fourth constant production factor – land – that is not subject to accumulation and depreciation. This

[7] See footnote 5 for our empirical definition of constancy. The longest series (fourty years) displaying a constant birth rate is given by the UK–data.

may imply that even the density of the population is relevant in determining development and that population growth exhibits diminishing returns in the long run. Finally, since physical and human capital accumulation of a country are sometimes strongly encouraged by foreign investments, future research on the topic may allow for a varying national savings rate. These external adjustments toward an efficient international division of labor may account for the remaining unobserved variation in our regressions.

In summary, the empirical evaluations of the last two chapters reported a significant positive short–run effect of GDP per capita growth on the variable birth rate, a significant positive–short run effect of death rate on birth rate and a significant negative long–run effect of birth rate changes on GDP per capita growth. All these results point toward the correctness of the four principles of classical economics and support the valid use of the classical unified growth model, which has moreover been found to be able to illustrate the stylized facts of a transition from an economic regime of stagnation, exhibiting high rates of births and deaths, to an economic regime of development, exhibiting steadily decreasing rates of births and deaths. Accepting these results as sufficiently mirroring reality, we will in the last chapter use the classical unified growth model to summarize and discuss its implications and to speculate about the future economic development suggested by this model.

Part IV

Implications of Classical Unified Growth Theory

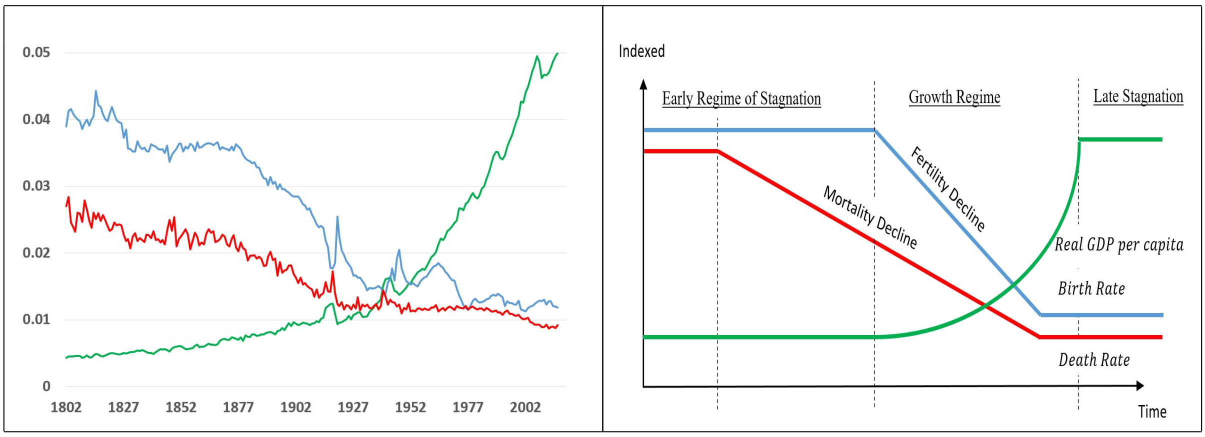
Chapter 9

Speculations on a Future Regime of Stagnation

9.1 The Recent Development

To assess the potential future economic development, we will continue the above line of thought that an early regime of stagnation is, in every economy, followed by a regime of development as long as mortality decreases. The left graph of Figure 9.1 extends the time horizon of Figure 6.1 by adding the most recent years of British history. Firstly, it is worth remarking that the death rate had achieved a low sustainable level of 0.125 as far back as the 1920s and has only slightly decreased since. Secondly, the birth rate roughly adjusted to the death rate during the 1970s and seems to have settled on a constant low level close to, but higher than the death rate. For better visualization, the presumed demographic transition from a weakly preventively checked toward a strongly preventively checked economy is stylized in the right graph of Figure 9.1.

Figure 9.1: Left graph: British "return to stagnation": Birth rate (blue), death rate (red) and GDP per capita (green) 1802–2016. Right graph: Stylized fact "return to stagnation".



Sources: GDP: Clark (2009) for 1800–1871, Mitchell (2013) for 1871–2010, Vital Rates: Wrigley and Schofield (1981) for 1800–1871, Mitchell (2013) for 1871–2010.

Thirdly, British real GDP per capita has slowed down to near–stagnation since the year 2007. Projecting this development into the future decades might suggest an additional stylized fact which remains to be evaluated, namely the return to what has lately been termed "secular stagnation", which is illustrated by the last regime "late stagnation" in the right graph of Figure 9.1. This idea of a new era of stagnation is sustained when extending the simulation by another one hundred periods, supposing a constant death rate (see Figures 9.2 and 9.3). We should, however, keep in mind that our simulation used a single labor division lag of merely fifteen years, whereas our estimations suggested a continuous labor division lag of approximately 65 years. The next section will give a short outline of the effects, the decisive operation of the great preventive check "principle of generation" is supposed to have on future population growth and growth in production according to our classical unified growth theory.

Figure 9.2: A simulation of a constant future death ratein growth rates (birth rate (blue), death rate (red), GDP per capita growth rate (light green)).

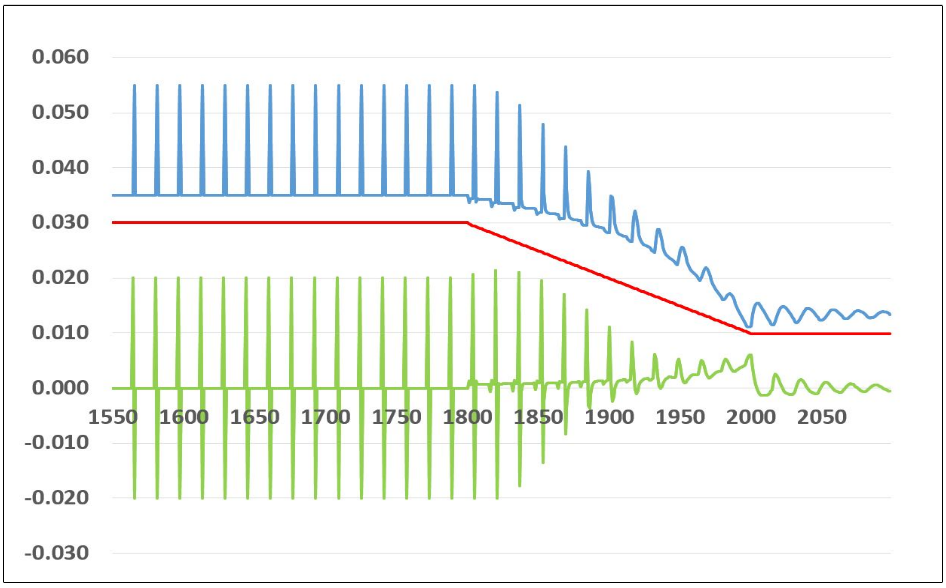
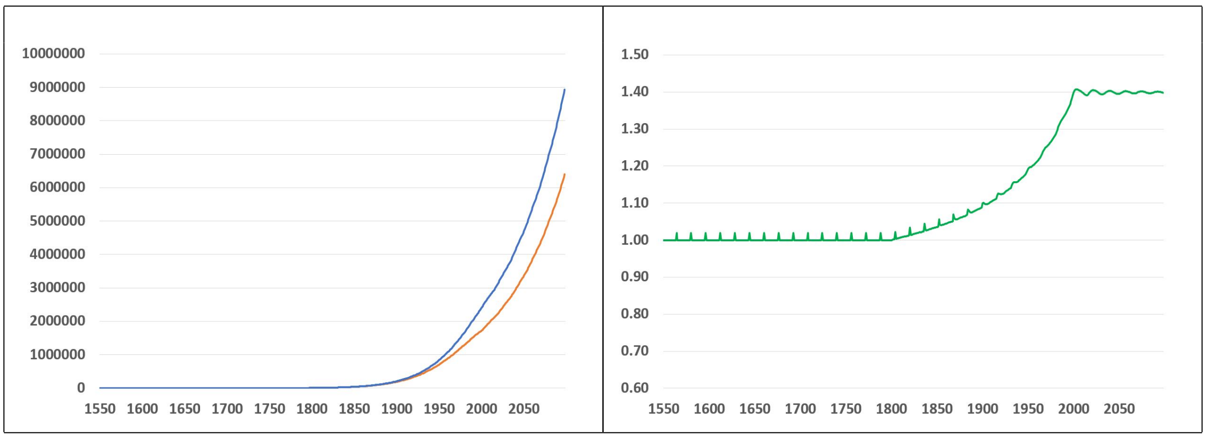


Figure 9.3: A simulation of a constant future death rate in level variables (population (orange), production (blue) and GDP per capita (green)).



9.2 Prognosis

Throughout this work, we used the simplifying assumption that the death rate cannot be affected endogenously. More generally, depending on "luck", it may increase or decrease, for example due to a potential future spread of new epidemics or new discoveries extending the maximum life expectancy. Based on the perhaps most realistic presumption of constant mortality, the classical unified growth model would suggest the following mechanism.

In the currently prevailing situation of low mortality in many economies in which the great preventive check operates extraordinarily strongly, the principle of generation would keep the birth rate on a constant low level as well. It would be concluded that approximately sixty years after the birth rate has adjusted to a lower level, an economy will enter a *new regime of stagnation*, where productivity stabilizes on a higher level. However, so far it is unclear from the data whether the birth rate will settle above, below or on the same level as the death rate, suggesting a corresponding positive, negative or zero future population growth rate. We will now briefly investigate these three possibilities as well as their outcomes suggested by the classical theory of population. As for the theory of production, it will also be asked how this change in population growth is expected to affect the future production and productivity of such an economy.[1]

Firstly, we assume that the birth rate would fall below the death rate. In this case, the cohort size of the descendants is even smaller than the cohort size of the ancestors. Since young individuals were neither threatened with being pushed out of existing niches, nor, owing to a decreasing population, facing potential economies of scale, niches would be ultimately passed on from ancestor to descendant without raising conflicts of interest. McCulloch even went a step further, believing that in an economy where the pressure of population and (intra- as well as intergenerational) competition would be eliminated, the corresponding "motives that stimulate intelligence and activity" would vanish as well. As a result, productive efforts, except for what is required to practice the parental profession, will become useless.

> If, indeed, it were possible that the stimulus arising from this principle [of population] would be suddenly removed, it is not easy to determine what life would be except a dreary blank, or the world except an uncultivated waste. Every exertion to which civilisation can be traced, proceeds, directly or indirectly, from its effects; either from the actual desire of having a family, or the pressing obligation of providing for one, or from the necessity of rivalling the efforts produced by the operation of these motives in others.[2]

Hence, productivity would ultimately stay unchanged while production would even fall in such an economy.

[1] Here, it might be helpful to remind the reader that the Solow model suggests that the size of population does not affect productivity. Independently of whether the population size was one billion, ten billion or twenty billion, productivity would stay unchanged over the long run, as it only reacts to *changes* in population growth. One might also generally argue that with a larger population some goods would become more expensive due to their scarcity, reflecting diminishing returns to population, while this loss would be neutralized by the larger supply of other goods that could be produced more cheaply due to an increasing division of labor and concomitant specialization.

[2] McCulloch (1863), part I, chapter VIII.

However, the principle of population also advises us to argue that the birth rate will – in the long run – not fall below the death rate. Since every deceasing individual leaves on average a vacant profession, the latter will tend to be occupied by an individual of an emerging cohort, supplying in turn the means of subsistence for his progeny. Consequently, since we expect a constant tendency of every population to at least replace itself, population growth will most arguably not become negative over the long–run and we reject the assumption of the birth rate falling below the death rate as being implausible. If we assume that the birth rate would remain on the same level as the death rate, actual population growth would be close to zero in the new regime. Since it seems obvious that a stagnating population will neither benefit from additional labor division nor lose from diminishing returns, production as well as productivity are of course equally expected to stagnate over the long run.

Finally, we advance the idea that the birth rate will stay above the death rate over the long run, with the principle of population sligthly outweighing the principle of generation. On the one hand, this would, as in the other two cases, lead to the – perhaps disillusioning – deduction that productivity would remain constant over the long run and it may even be the case that this new regime of stagnation will endure for another millenium. On the other hand, the notion that population pressure triggers innovations may at least give us some hope of continued "technological progress". In the end, the principle of population remains the main agent that drives a population into competition, making a division of labor and specialization ultimately inevitable, since specialization does not emerge in the absence of population pressure. Thus, although it was primarily perceived by many economists as a source of misery, and less as a driver of total economic output, the impact of the principle of population on economic growth is perhaps best summarized by the final appraisal the principle received in classical economics:

> The principle of increase, as explained by Malthus [1798], [. . .] appeared to form an insuperable obstacle to all permanent improvement in the condition of society, and to condemn the great majority of the human race to a state approaching to destitution. But farther inquiries have shown that the inferences drawn [. . .] from the principle [. . .], are contradicted by the widest experience; that the too rapid increase of population is almost always prevented by the influence of principles which its increase brings into activity; that a vast improvement has taken place in the condition of the people of most countries [. . .] and that, so far from being inimical to improvement, we are really indebted to the principle of increase for most part of our comforts and enjoyments, and for the continued progress of arts and industry. [. . .] That the tendency to increase is not inconsistent with the improvement of society, is a fact as to which there can be no dispute.[3]

Regardless of whether specialization and innovation are caused by "want" or "the fear of losing a social rank", they are the regular result of the current cohort outnumbering the former, generating intergenerational competition and driving an abundant number of young individuals into new niches. If in addition, inspite of population growth, the average number of abundant descendants becomes sufficiently small to be constantly absorbed by new forms of specialization, the demand for which arose from the increased market size, there exists a high probability that the resulting new professions betimes allowed for mortality–reducing innovations.

[3] McCulloch (1863), preface.

Chapter 10

Conclusion

10.1 Recapitulation

The aim of this work was to construct a unified theory of economic growth that can explain the stylized evolution of GDP per capita over the very long run, i.e. over the complete history of mankind, in a single coherent model. In the introductory chapter, the evolution of GDP per capita was illustrated and the two most prominent unified growth models were presented. While the older model by Kremer postulates one continuous process of economic development, the newer model by Galor and others suggests that the process of economic growth can be subdivided into two regimes – a regime of economic stagnation and a regime of economic development.

In chapter two, we found empirical support for Galor's historical regime of stagnation in GDP per capita in Britain until around 1800 AD. Since Galor's unified growth model is chiefly based on a mechanism that had already been described by the classical economist T. R. Malthus in the year 1798, he employed the term "the Malthusian trap." In chapter three, this mechanism was traced back to Malthus' and other classical economists' original understanding, which made use of three universal principles. On this basis, a mathematical model of the "classical Malthusian trap" relying on these principles was built and its plausibility was illustrated.

In chapter four, the stylized facts of the (second) regime of economic development were analyzed and the classical model was used to explain this process of development. Here, it was also shown that the most popular contemporary approaches cannot explain the "escape from the Malthusian trap." As a promising alternative, this work advanced a fourth classical principle in chapter five, by which the switch from a regime of stagnation to a regime of development could be properly explicated. In chapter six, the mechanism of the Malthusian trap as well as the escape from the Malthusian trap were simulated using one and the same model to account for a unified growth theory.

In chapter seven and chapter eight, the four classical principles on which the classical unified growth theory rests were empirically evaluated using OLS regressions. Chapter nine concluded by speculating about future economic growth and development by using the findings from classical unified growth theory.

10.2 Summary of the Main Findings

To provide an overview of the most important findings in this work, this section will list the main achievements following the structure of the chapters. Firstly, chapter two found considerable evidence of the existence of a regime of economic stagnation in Britain until 1800 AD and as a global phenomenon on a national level until today. Secondly, we found that the stylized fact of a cycle of misery between population and production can be explained by advancing a classical theory of production and a classical theory of population, consisting of the principle of labor division, the principle of diminishing returns and the principle of population, conjointly forming an endogenous demographic growth mechanism, the Malthusian trap. Since this mechanism is still not universally accepted in the science of economics, chapter two and chapter three mainly served to clarify its importance and to lay a firm ground for investigations on the regime of economic development. The best proof of the existence of the Malthusian trap is provided by the observed confirmation of the Darwinian theory of evolution by natural selection.

Thirdly, with regard to the regime of economic development, it has been found that a population slowdown must have been responsible for the breakout from stagnation. This breakout was not caused by the industrial revolution as an agent of technological progress. Instead, it was caused by a fertility decline, which Malthus had predicted to reflect the "great preventive check." Fourthly, and this should be considered as the one most crucial finding of this work, an extension of the classical growth model using the great preventive check "principle of generation" as fourth universal principle is able to explain the fertility decline through decreasing mortality. Accordingly, the fertility decline was not caused by a reversal of the principle of population, as is often falsely asserted. Fifthly, we found evidence of an "exogenous" mortality decline. As a result, by simulating the extended classical growth theory it was shown that the model can account for the three combined stylized facts "cross of wealth" (declining fertility and increasing productivity), "demographic transition" (declining mortality and declining fertility) and "epidemiological transition" (declining mortality due to the decreased prevalence of infectious diseases) by advancing a simple chain of causality: Declining mortality causes fertility to decline, which enables productivity to increase. This theory is in line with the finding that there exists not a single economy in which a long–run increase in GDP per capita can be observed without a simultaneous decline in fertility and a preceding decline in mortality.

The narrative of the main finding of this work, the great preventive check principle of generation, is as follows: As the positive checks to population (famines, epidemics, wars etc.) are supposed to be non–existent when the decline in mortality is complete, the question arose of how Malthus' "obvious and powerful" preventive checks could be defined which are claimed to be capable of reducing the rate of population growth within manageable limits? To this end, it has been noted that declining mortality implies (almost by definition) rising life expectancy. As the average lifetime of a parent increased, his average working time was extended as well. Property rights, inheritance, experience and reputation served the older cohorts to secure their economic niche against upcoming competition from the younger cohorts. With a limited amount of niches available, a large fraction of the young generation would find it necessary to comply with the advice not to procreate until they are established in a profession which may enable them to support a family without fearing the loss of their social rank. This generation conflict induces delayed marriages and declining fertility. Consequently, it has been argued that chiefly a decrease in mortality ought to be aimed at when trying to promote economic development. Moreover, this process of intergenerational competition has been found to be in accordance with the Darwinian theory of evolution by sexual selection.

Sixthly, the advanced classical growth theory has been shown to be capable of modeling a regime of stagnation and a regime of development using one and the same unified growth model. Seventhly, since we had so far merely reconstructed the stylized facts of stagnation and development as a result of the concerted action of the four principles, it was considered necessary to evaluate the effects of the four principles separately. On the one hand, we found evidence of the principle of population and the principle of generation by using around 4,500 observations of annual international data. A vector autoregression allowed us to conclude that there exists an almost universal positive accumulated (short–run) effect of GDP per capita growth on birth rates as well as an equal effect of death rates on birth rates after four years. On the other hand, conjoint evidence of the principle of diminishing returns and the principle of labor division was confirmed by applying OLS regressions from data of 104 countries to measure the average (long–run) effect of a change in birth rates on a GDP per capita growth. The results suggested that a 1% decrease in birth rate induces GDP per capita growth to increase by 2%, spread over the subsequent 65 years. The robustness of the principles has been confirmed by investigating country–specific national data.

Eightly, when using the classical unified growth model as a forecast, it points toward a future regime of economic stagnation in GDP per capita as long as mortality remains constant. Nonetheless, since the birth rate is expected to stay above the level of the death rate, it is assumed that this regime of stagnation is accompanied – as in the case of the early regime of stagnation – by population growth and "technological progress."

10.3 Consequences for the History of Economic Thought

As every student of the history of economic thought knows, Malthus suggested that population had an inherent tendency to grow, that it would grow whenever the means of subsistence increased and that it would even tend to outgrow the means of subsistence. Thanks to Darwin, we now know that this tendency resulted in competition and a struggle for life that has been fundamental to the origin of species. Moreover, the classical economists – in most cases well educated economic historians – understood that the doctrine of competition constituted the foundation of the Smithian division of labor and thus of their entire theory of economic growth. Accordingly, when intending to tread in their footsteps, it is essential to acknowledge that the general tendency for competition is a result of a universally operating principle of population and that nothing makes sense in classical economic growth theory if not seen in the light of the theory of population.[1]

[1] The reason for mainly relying on the logical arguments of the classical economists in this work may be due to the zeitgeist of the nineteenth–century, when deep and profound theoretical thinking was much more

Notwithstanding its well-founded theoretical and empirical reasoning, there remains a fair portion of skeptics, some of whom seem to have interpreted the classical theory of population surprisingly incorrectly. Although these authors frequently refer to Malthus (1798) as chief historical source and economic authority, it seems that neither the critics nor the majority of the proponents of the classical theory of population have consulted Malthus' later editions (1803-1826) on the principle of population. If they had done so, they might have arrived at the insight that the proposed theory not only supplies a mechanism of a stylized historical regime of stagnation, but at the same time offers a mechanism by which stagnation could be overcome. Malthus saw the possibility that productivity could increase and that fertility could decrease in the long run – a prediction that has become reality in all developed countries. Moreover, he reasoned that if fertility decreased, then productivity must rise, and that declining fertility was the only way to escape the population trap. Most importantly, he suggested how fertility was to be decreased, namely by reducing mortality. In this sense, Keynes' famous assessment may too be best understood:

> If only Malthus, instead of Ricardo, had been the parent stem from which nineteenth-century economics proceeded, what a much wiser and richer place the world would be today! We have laboriously to rediscover and force through the obscuring envelopes of our misguided education, what should never have ceased to be obvious.[2]

Hence, this work recommends closing the chapter of dispelling Malthusian ideas over the past two hundred years. Nonetheless, the author is quite aware that even if economists were to strongly support the advancement of the Malthusian theory, it is questionable whether it will ever become generally accepted.[3]

10.4 Consequences for Economic History

Likewise, economic historians should be aware that the Malthusian effects have always been well in place and persist until today. Moreover, it is noteworthy that the industrial revolution and technological progress were often not by themselves capable of increasing average economic productivity, albeit increasing economic output. Thirdly, we should bear in mind that, as a consequence of the tendency to increase, human history and evolution have been constantly accompanied by population expansion as follows. The rise and fall of civilizations may often be ascribed to the degree of population growth. If the environment allowed for a strong increase, we tend to observe urbanization, specialization and the rise of a culture. However, as soon as the economy has reached its natural boundaries and therefore ceased to be capable of expansion, the principle of generation tends to regulate the increase in numbers, often resulting in stagnation. Stagnation, on the other hand, has in many historical instances proved lethal, particularly if nearby civilizations continued to expand in numbers at the same time.

appreciated, as it was harder to discover correlations between variables without having computer software and big data at hand.

2 Keynes (1933), Essays in Biography, Malthus, pp. 120–121.

3 "[W]hat a discouraging example Malthus is, to show during what long years the plainest case may be misrepresented and misunderstood." Darwin (1860).

10.5 Consequences for Economic Growth Theory

The purpose of this work was to provide and validate a theory that solves the economic problem of stagnation and growth, or, in other words, to disentangle the effects responsible for a historical regime of economic stagnation and for a regime of economic development. Hitherto, the field of unified growth theory has attempted to offer a theoretical analysis of the relationship between the demographic transition and the economic transition to development. Having retraced unified growth theory to its classical demographic predecessor, four classical elementary principles were interpreted to account for the relevant interactions between demographic and economic variables. While the principle of diminishing returns and the principle of labor division are commonly acknowledged in economic theory in one form or another, the existence of the principle of population remains debated. Furthermore, the great preventive check, the principle of generation, has even been ignored in most arguably all recent evaluations of the Malthusian model. However, when accounting for the last two principles, classical theory is found to match the stylized facts of stagnation and development. These four universal principles should be accepted as constantly operating economic "laws", although the displayed effects of the principle of population and the principle of generation are observed to vary over time. The universality of the principle of population must be particularly emphasized, as it seems that its effect has been regularly misinterpreted in the past.

The resulting endogenous theory of growth can be subdivided into a theory of production and a theory of population. The advanced classical theory of production is largely in line with the neoclassical Solow model – one the best known and probably most frequently used models in the field of economics. Here, growth economists are advised to recognize that the simple Solow model without technology is based on the classical principles of diminishing returns and labor division, exhaustively defining the classical theory of production by employing the axioms of constant returns to scale and that population is the source of all value. The corresponding steady–state equation suggests that, if positive population growth is reduced from a high level to a lower level, population pressure from diminishing returns relaxes while the earlier established division of labor derived from a formerly higher population growth continues to have a positive effect. Hence, the findings of the theory of production advanced in this work are easily summarized: A change in population size affects production proportionally over the long–run, while a change in population growth produces a negative reaction of production per capita. With regard to the theory of production, growth economists are therefore advised to continue using the Solow model without technological progress as a proper tool of modeling economic growth.

In addition, since the theory of production assumed exogenous population growth, an endogenous growth theory required the knowledge of the determinants of population growth, or in other words, a theory of population. This work advanced the view that the reference point in economic–demographic investigations should be the state of a Malthusian trap, where population growth steadily tends to outperform economic growth. Having agreed that population has the potential to outgrow GDP, we next asked what factors determined the speed of population growth. Firstly, contrary to what has recently been implicitly assumed by a majority of growth economists, there exists a difference between the classical principle of population and the Malthusian trap. The Malthusian trap defines a state of stagnating economic productivity resulting from the operation of an unrestricted principle of population. As the Malthusian trap is a *testable fact*, its existence was – following Malthus' experience – first reasonably verified and later equally reasonably falsified. In contrast, the principle of population is the incontrovertible *tendency* of each living being to increase in numbers whenever its means of subsistence increase. As the second determinant of population growth, growth economists are advised to model the great preventive check principle of generation, which can account for the escape from the Malthusian trap. Although the principle of generation has been extensively outlined in this work, it cannot be said to have been exhaustively determined, except for stating that it is triggered by diminishing mortality.

Combining the theory of population and the theory of production, we may state that if the principle of population is allowed to operate freely, it causes rapid population growth, strong competition for resources and consequently hardship and innovation. If, on the other hand, population pressure was fully suppressed preventively, lack of competition would diminish the number of innovations to a minimum level. As the two extremes of unchecked and fully checked population growth are obvious, it follows that there must be some transitional, intermediate point in which the two principles outweigh each other to such a moderate degree as to allow for a slow increase of population and at the same time for an equally moderate pursuit of innovation such that economic development is optimized.

To pursue the classical view, the above findings recommend a further move away from exogenous toward endogenous growth theories and from short–run theories toward unified growth theories to explain economic development over the very long run in one framework together with the demographic transition.

10.6 Consequences for Economic Development

To trigger the transition to economic development, classical unified growth theory proposes to reduce mortality or, what is almost the same, to increase life expectancy. The demographic structure resulting from such a change is much less prone to overpopulation, as a major part of the population, in particular the part exhibiting a high social status, becomes infertile. If it would not have been for the principle of generation, the population of the earth would certainly not have been confined to less than ten billion inhabitants. As a practical result, classical unified growth theory asserts that the increase in production per capita over the past two hundred years was chiefly owed to the epidemiological transition. Employing descriptive statistics, we have illustrated for a large body of countries that there was indeed an epidemiological transition followed by a demographic transition, which is in some parts of the world not yet finished and that incomes were low before and increased strongly with the onset of the demographic transition. While the size of a population has been found to have a neutral effect on economic development, our theory also suggests that a population growth rate *change* (in this framework measured in terms of birth rate changes) may be the best predictor for economic development. Since developmental economists currently tend to investigate the causes for improvements in GDP per capita by conducting regressions on an almost infinite number of – often relatively arbitrary – variables, it may be advisable to return to growth regressions with a sound theoretical fundament.

Although this work does not explicitly account for political interventions, we will shortly consider in how far government action might speed up or slow down the process of development. Firstly, as a consequence of the principle of generation, the most practicable and probably most human way to limit population pressure lies in a reduction of the death rate that is largely kept high by epidemics such as currently in sub–saharan regions. This reduction of mortality will certainly raise the population pressure on the upcoming generation for some time and correspondingly increase poverty. However, so far no instance has been observed in which decreasing fertility and increasing productivity was not preceded by such a transitional period. Secondly, with the simultaneous establishment of the principle of population and the principle of generation, classical unified growth theory yields an explanation that can solve the "demographic economic paradox", which stated that economies with higher productivity tend to exhibit lower birth rates. For policy implications, it is important to realize that – in accordance with our empirical findings – there exists no such negative causal link running from living standards to fertility. On the contrary, development support in the form of wealth might even aggravate the population problem, since we observe a significant and robust positive effect of growth in living standards on fertility. Bluntly stated, "development is not the best contraceptive", but "contraception is the best development."

10.7 Consequences for "Natural Philosophy"

In this dissertation, I have tried to clarify the theoretical channels through which economic development works. There is no reason to doubt that an increase in life expectancy such that two productive generations exist at the same time will (ceteris paribus) decisively delay the propagation of its offspring and therefore reduce fertility. We will find the principle of population and the principle of generation to prevail in every proper experiment on any natural species. If, notwithstanding its evidence, we questioned or even rejected the most obvious channels of operation – inheritance, infant mortality, sexual selection and infertility – we ought to find some other answer that explains why higher life expectancy reduces fertility.

Individual "enlightenment" has only very recently been perceived as an important driver of economic development due to the fact that health has increased considerably and the average life span of individuals has doubled over the past two hundred years, therefore increasing the ability of the average individual to accumulate human capital in the form of education and experience. On the contrary, in spite of what is often denoted as prudence, foresight or reason, the spirit of materialism present throughout this work assumed that, although human individuals undoubtedly possess a higher potential for abstract thinking, i.e. more "rational" expectations, than other living beings, they do not in reality decide self–reliably on the number of their progeny, but they react to existing circumstances.[4] The classical economists largely argued that individual improvement would have ultimately been in vain, as long as population continued to

[4] the existence of the Malthusian trap up until around 1800 AD - and in many countries until today - proves that human communities are often not willing to collectively restrict their numbers. How else can we explain

increase.[5] This materialistic view enabled – albeit this happened at the cost of a very high level of abstraction – the classical economists to model the human economy in the same simple and parsimonious manner as any animal economy by merely considering the variables birth rate, death rate and productivity as exhaustively causing the development of the ratio between production and population over the long run.

Accordingly, the history of mankind should be viewed neither as a history of technological progress, nor as a history of political or institutional progress and much less as a history of continuously improving individual enlightenment. The history of mankind is a history of population and the principle of population remains the "big dark cloud above the enlightenment movement." Since population is the only variable that possesses an inherent tendency to grow, it causes at best a division of labor and what is generally perceived as "progress" in the form of technological and institutional "improvement". Whether an additional person contributes positively to average economic productivity seems – although in the short run influenced by the way education has been conducted and by contemporaneous economic circumstances – to depend on an evolutionary process of trial and error. This process cannot be imagined to take place without population pressure.

If the above theory is independently confirmed, it ought to be ranked as one of the most illuminating achievements of thought on which every social science should be built upon. However, in spite of this bold statement,

> I by no means expect to convince experienced [economists] whose minds are stocked with a multitude of facts all viewed, during a long course of years, from a point of view directly opposite to mine. It is so easy to hide our ignorance under such expressions as ["technology", "institutions", "enlightenment"] etc., and to think that we give an explanation when we only restate a fact. Any one whose disposition leads him to attach more weight to unexplained difficulties than to the explanation of a certain number of facts will certainly reject the theory. [...]; but I look with confidence to the future, to young and rising [economists], who will be able to view both sides of the question with impartiality.[6]

the fact that immigrants tend to relatively quickly adjust to the habits of pairing and childbearing prevailing in the new country?

[5] "The purposive adjustment of the birthrate to the supply of the material potentialities of well-being is an indispensable condition of human life and action, of civilization, and of any improvement in wealth and welfare. [...] social cooperation is impossible if people give rein to the natural impulses of proliferation." v. Mises (1949).

[6] Darwin (1859), conclusion.

Bibliography

Abramovitz, Moses. 1956. "Resource and Output Trends in the United States Since 1870." In Resource and Output Trends in the United States Since 1870, 1–23. National Bureau of Economic Research, Inc.

Acemoglu, Daron, and James A. Robinson. 2012. Why Nations Fail. 1st ed. New York: Crown Publishers.

Alison, William Pulteney. 1847. Observations on the Famine of 1846-7, in the Highlands of Scotland and in Ireland. Edinburgh: W. Blackwood and Sons.

Allen, Robert C. 2001. "The Great Divergence in European Wages and Prices from the Middle Ages to the First World War." Explorations in Economic History, 411–447.

—. 2009. The British Industrial Revolution in Global Perspective. Cambridge: Cambridge University Press.

Alter, George, and Gregory Clark. 2010. "The Demographic Transition and Human Capital." In The Cambridge Economic History of Modern Europe, Volume 1, 1700-1870, by S. N. Broadberry and Kevin H. O'Rourke, 43–69. Cambridge: Cambridge University Press.

Angeles, Luis. 2010. "Demographic Transitions: Analyzing the Effects of Mortality on Fertility." Journal of Population Economics, 99–120.

Artrzrouni, Marc, and John Komlos. 1985. "Population Growth Through History and the Escape from the Malthusian Trap: A Homeostatic Simulation Model." Genus, 21–39.

Ashraf, Quamrul H., David N. Weil, and Joshua Wilde. 2013. "The Effect of Fertility Reduction on Economic Growth." Population and development review, 97–130.

Bairoch, Paul. 1988. Cities and Economic Development from the Dawn of History to the Present. 1988. Chicago: University of Chicago.

Barro, Robert J., and Xavier Sala-i-Martin. 2003. Economic Growth. Prentice Hall India.

Bastiat, Frederic. 1860. "Harmonies of Political Economy." Patrick S. Sterling. London: J. Murray.

Becker, Gary S., and Gregg H. Lewis. 1973. "On the Interaction Between the Quantity and Quality of Children." Journal of Political Economy, 279–288.

Becker, Gary Stanley. 1981. A Treatise on the Family. Cambridge (Massachusetts): Harvard University Press.

—. 1988. "Family Economics and Macro Behavior." American Economic Review, 1–13.

Becker, Gary Stanley, and K. M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." The Quarterly Journal of Economics, 1137–1160.

Bloom, David E., David Canning, and Jaypee Sevilla. 2003. The Demographic Dividend. Santa Monica, Calif.: Rand.

Boserup, Ester. 1965. The Conditions of Agricultural Growth. London; Chicago: Allen & Unwin.

Broadberry, Stephen, and Kevin H. ORourke. 2010. The Cambridge Economic History of Modern Europe. Cambridge: Cambridge University Press.

Broadberry, Stephen, Bruce M. S. Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. 2015. British Economic Growth, 1270–1870. Cambridge: Cambridge University Press.

Burnette, Joyce. 2006. "How Skilled Were English Agricultural Labourers in the Early Nineteenth Century?" The Economic History Review, 688–716.

Carey, Henry Charles. 1837. Principles of Political Economy. Vol. I. Philadelphia: Carey, Lea & Blanchard.

Cervellati, Matteo, Uwe Sunde, and Klaus F. Zimmermann. 2017a. "Demographic Dynamics and Long-Run Development: Insights for the Secular Stagnation Debate." Journal of Population Economics, 401–432.

—. 2017b- "Demographic Change and Long–Run Development." CESifo Seminar Series. The MIT Press. Cambridge, Massachusetts. London, England.

Chatterjee, Shoumitro, and Tom Vogl. 2018. "Escaping Malthus: Economic Growth and Fertility Change in the Developing World." American Economic Review, 1440–1467.

Cipolla, Carlo M., J. C. Russell, R. J. Mols, and A. Armengaud, . 1971. Bevölkerungsgeschichte Europas. Vol. 19. München: Piper.

Clark, Gregory. 2007. A Farewell to Alms. Princeton University Press.

—. 2008. "In Defense of the Malthusian Interpretation of History." European Review of Economic History, 175–199.

—. 2018. "Growth or stagnation? Farming in England, 1200-1800." The Economic History Review, 55–81.

—. 2009. "The Macroeconomic Aggregates for England, 1209–2008." Vol. 27, in Research in Economic History, by Alexander J. Field, 51–140. Bingley, U.K.: Emerald.

Clark, John Bates. 1907. Essentials of Economic Theory. New York: The Macmillan COmpany.

—. 1899. The Distribution of Wealth: a Theory of Wages, Interest and Profit. New York: Macmillan.

Cobb, Charles W., and Paul H. Douglas. 1928. "A Theory of Production." The American Economic Review, 139–165.

Crafts, Nicholas, and Terence C. Mills. 2009. "From Malthus to Solow: How did the Malthusian economy really evolve?" Journal of Macroeconomics, 68–93.

Darwin, Charles. 1871. The Descent of Man, and Selection in Relation to Sex. 1. Vol. 1. London: John Murray.

—. 1859. The Origin of Species: By Means of Natural Selection Or the Preservation of Favored Races in the Struggle for Life. 2. New York: D. Appleton and Company. 2nd edition 1872.

—. 1868. The Variation of Animals and Plants under Domestication. 1. London: John Murray.

Darwin, Charles. 1993. "To Charles Lyell, 8th June [1860]." In The Correspondence of Charles Darwin: Volume 8, 1860, by Charles Darwin, Frederick Burkhardt and Sydney Smith. Cambridge: Cambridge University Press.

Davenport, Romola, Leonard Schwarz, and Jeremy Boulton. 2011. "The Decline of Adult Smallpox in Eighteenth-Century London." The Economic History Review, 1289–1314.

Diamond, Jared M. 2005. Collapse. 2. New York: Penguin Books.

Domar, Evsey. 1946. "Capital Expansion, Rate of Growth, and Employment." Econometrica, 137.

Easterly, William. 2001. The Elusive Quest for Growth. Cambridge, Mass.; London: MIT Press.

Eckstein, Zvi, T.Paul Schultz, and Kenneth I. Wolpin. 1986. "Short-Run Fluctuations in Fertility and Mortality in Pre-Industrial Sweden." European Economic Review, 295–317.

Edvinsson, Rodney Benjamin. 2017. "The Response of Vital Rates to Harvest Fluctuations in Pre-Industrial Sweden." Cliometrica, 245–268.

Ehrlich, Paul Ralph. 1968. The Population Bomb. 1. New York: Ballantine Books.

Fanchon, Phillip, and Jeanne Wendel. 1992. "Estimating VAR Models under Non-Stationarity and Cointegration: Alternative Approaches for Forecasting Cattle Prices." Applied Economics, 207–217.

Fernihough, Alan. 2012. "Malthusian Dynamics in a Diverging Europe: Northern Italy, 1650-1881." Demography, 311–332.

Foreman-Peck, James. 2019. "Economic-Demographic Interactions in the European Long Run Growth." In Handbook of Cliometrics, by Claude Diebolt and Michael Haupert, 1–25. Berlin, Heidelberg: Springer Berlin Heidelberg.

Galor, Oded. 2011. Unified Growth Theory. Princeton: Princeton University Press.

Galor, Oded, and David N. Weil. 1999. "From Malthusian Stagnation to Modern Growth." American Economic Review, 150–154.

Haines, Michael R. 1998. "The Relationshp Between Infant and Child Mortality and Fertility: Some Historical and Contemporary Evidence for the United States." In From Death to Birth : Mortality Decline and Reproductive Change, by Mark R. Montgomery and Barney Cohen, 227–253. Washington (DC).

Hajnal, John. 1965. "European Marriage Patterns in Perspective." In Population in History. Essays in Historical Demography, by D. V. Glass and D. E. C. Eversley, 101–143. New Brunswick: Aldine Publishing Company.

Hansen, Gary D., and Edward C. Prescott. 2002. "Malthus to Solow." American Economic Review, 1205–1217.

Hardin, Garett. 1999. Garrett Hardin on Malthus, Darwin and Natural Selection Educational Communications, Inc. https://www.youtube.com/watch?v=0NdHfB2tFwM.

Hardin, Garrett. 1968. "The Tragedy of the Commons." Science (New York, N.Y.), 1243–1248.

Harrod, R. F. 1939. "An Essay in Dynamic Theory." The Economic Journal, 14.

Hays, J. N. 2005. Epidemics and Pandemics. Santa Barbara, Calif.; Oxford: ABC-CLIO.

Herzer, Dierk, Holger Strulik, and Sebastian Vollmer. 2012. "The Long-Run Determinants of Fertility: One Century of Demographic Change 1900–1999." Journal of Economic Growth, 357–385.

Hicks, J. R. 1936. "Distribution and Economic Progress: A Revised Version." The Review of Economic Studies, 1.

Hinde, Peter, Gregori Galofré-Vilà, and Aravinda Meera Guntupalli. 2018. "Heights Across the Last 2000 Years in England." Research in Economic History, 67–98.

Humphrey, Thomas M. 1997. "Algebraic Production Functions and Their Uses before Cobb-Douglas." Federal Reserve Bank of Richmond Economic Quarterly, 51–83.

Huxley, Julian S. 1926. Essays in Popular Science. Knopf, A. A.

Kalemli-Ozcan, Sebnem. 2002. "Does the Mortality Decline Promote Economic Growth?" Journal of Economic Growth, 411–439.

Kalemli-Ozcan, Sebnem, and David N. Weil. 2010. "Mortality Change, the Uncertainty Effect, and Retirement." Journal of Economic Growth.

Kelly, M. 1997. "The Dynamics of Smithian Growth." The Quarterly Journal of Economics, 939–964.

Keynes, John Maynard. 1930. "Economic Possibilities for Our Grandchildren." In Essays in Persuasion, by John Maynard Keynes, 358–373. New York: Hartcourt Brace.

—. 1933. Essays in Biography. London: Macmillan and Co.

—. 1919. The Economic Consequences of the Peace. 1. London: Macmillan & Co., Limited.

—. 1936. "The General Theory of Employment." The Quarterly Journal of Economics, 209–223.

Kirk, D. 1996. "Demographic Transition Theory." Population Studies, 361–387.

Komlos, John. 2003. "The Industrial Revolution as the Escape from the Malthusian Trap." Journal of European Economic History, 307–331.

Korotayev, Andrey, Sergey Malkov, and Leonid Grinin. 2014. "A Trap at the Escape from the Trap? Some Demographic Structural Factors of Political Instability in Modernizing Social Systems." By Leonid Grinin and Andrey Korotayev. Volgograd: Uchitel Publishing House.

Kremer, M. 1993. "Population Growth and Technological Change: One Million B.C. to 1990." The Quarterly Journal of Economics, 681–716.

Lange, Jérôme. 2012. "Allyn Young's Inquiry into the Nature and Causes of the Growth of Markets." Working Paper Univerité Paris Descartes.

Langer, William L. 1963. "Europe's Initial Population Explosion." The American Historical Review, 1.

Lassalle, Ferdinand. 1863. Offenes Antwortschreiben an das Zentral-Komitee zur Berufung eines Allgemeinen Deutschen Arbeiter-Kongresses zu Leipzig. Friedrich Hertneck.

Lee, Ronald D. 1981. "Short-Term Variation: Vital Rates, Prices and Weather." In The Population History of England, 1541-1871, by Edward A. Wrigley and Roger S. Schofield, 356–401. London: Edward Arnold.

Lee, Ronald D., and Michael Anderson. 2002. "Malthus in State Space: Macro Economic-Demographic Relations in English History, 1540 to 1870." Journal of Population Economics, 195–220.

Liso, Nicola, Giovanni Filatrella, and Nick Weaver. 2001. "On Endogenous Growth and Increasing Returns: Modeling Learning-by-Doing and the Division of Labor." Journal of Economic Behavior & Organization, 39–55.

Livi-Bacci, Massimo. 2012. A Concise History of World Population. 5th ed. Oxford: Wiley-Blackwell.

Lucas, Robert E. 1988. "On the Mechanics of Economic Development." Journal of Monetary Economics, 3–42.

Maddison, Angus. 2006. The World Economy. Vol. 2. Paris: OECD Publishing.

Malthus, Thomas Robert. 1798. An Essay on the Principle of Population As It Affects the Future Improvement of Society. London: Reeves and Turner.

—. 1803. An Essay on the Principle of Population; Or, A View of Its Past and Present Effects on Human Happiness. 2. London: John Murray.

—. 1826. An Essay on the Principle of Population; Or, A View of Its Past and Present Effects on Human Happiness. 6. London: John Murray.

Mankiw, N. G., Romer, and D. N. Weil. 1992. "A Contribution to the Empirics of Economic Growth." The Quarterly Journal of Economics, 407–437.

Marshall, Alfred. 1890. Principles of economics: unabridged eighth edition 2009. Cosimo, Inc.

McCloskey, Deirdre N. 2010. Bourgeois Dignity: Why Economics can't Explain the Modern World. University of Chicago Press.

McCulloch, John Ramsay. 1863. A Treatise on the Principles and Practical Influence of Taxation and the Funding System. Edinburgh: Adam and Charles Black.

McKeown, Robert E. 2009. "The Epidemiologic Transition: Changing Patterns of Mortality and Population Dynamics." American Journal of Lifestyle Medicine, 19–26.

McKeown, Thomas. 1965. "Medicine and World Population." Journal of Chronic Diseases, 1067–1077.

Mill, John Stuart. 1848. The Principles of Political Economy with some of their Applications to Social Philosophy. London: John W. Parker.

Mises, Ludwig von. 1949. Human Action. Yale University Press.

Mitchell, Brian R., ed. 2013. International Historical Statistics, 1750-2010. London: Palgrave Macmillan.

Mokyr, Joel. 2016. A Culture of Growth. Princeton University Press.

Mokyr, Joel, and Hans-Joachim Voth. 2010. "Understanding Growth in Europe, 1700–1870: Theory and Evidence." In The Cambridge Economic History of Modern Europe, by Kevin H. O'Rourke and S. N. Broadberry, 7–42. Cambridge: Cambridge University Press.

Møller, Niels Framroze, and Paul Sharp. 2014. "Malthus in Cointegration Space: Evidence of a Post-Malthusian Pre-Industrial England." Journal of Economic Growth, 105–140.

Mun, Thomas. 1664. England's Treasure by Forraign Trade. London.

Nicolini, E. A. 2007. "Was Malthus Right? A VAR Analysis of Economic and Demographic Interactions in Pre-Industrial England." European Review of Economic History, 99–121.

North, Douglass C. 1992. Transaction Costs, Institutions, and Economic Performance. Vol. 30. San Francisco: ICS Press.

North, Gary. 2013. How Come We're So Rich? YouTube. Austrian Economics Research Conference. https://www.youtube.com/watch?v=wqmZXvDjEBA.

Nunn, Nathan, and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." Quarterly Journal of Economics, 593–650.

Pfister, Ulrich, and Georg Fertig. 2010. "The Population History of Germany: Research Strategy and Preliminary Results." no. WP-2010-035. Max Planck Institute for Demographic Research.

Rathke, Alexander, and Samad Sarferaz. 2014. "Malthus and the Industrial Revolution." no. 4667.

Robinson, Joan. 1954. "The Production Function and the Theory of Capital." The Review of Economic Studies, 81–106.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." Journal of Political Economy, 1002–1037.

Sachs, Jeffrey, and Pia Malaney. 2002. "The Economic and Social Burden of Malaria." Nature, 680–685.

Senior, Nassau William. 1836. An Outline of the Science of Political Economy. London: W. Clowes and sons.

Sims, Christopher A. 1980. "Macroeconomics and Reality." Econometrica, 1.

Smith, Adam. 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. London: W. Strahan, and T. Cadell.

Solow, Robert M. 2001. "After "Technical Progress and the Aggregate Production Function"." In New Developments in Productivity Analysis, 173–178. National Bureau of Economic Research, Inc.

—. 1956. "A Contribution to the Theory of Economic Growth." The Quarterly Journal of Economics, 65–94.

—. 1957. "Technical Change and the Aggregate Production Function." The Review of Economics and Statistics, 312–320.

Spencer, Herbert. 1874. Principles of Sociology in Three Volumes. D. Appleton and Company.

—. 1852. "A Theory of Population, deduced from the General Law of Animal Fertility." The Westminster Review, 468–501.

Swan, T. W. 1956. "Economic Growth and Capital Accumulation." Economic Record, 334–361.

Tanner, James M. 1994. "Introduction: Growth in Height as a Mirror of the Standard of Living." Stature, Living Standards and Economic Development: Essays in anthropometric history, London, 1–6. In Komlos, A. U. 1994. Stature, Living Standards, and Economic Development. Chicago: University of Chicago Press.

Theil, Henri. 1971. Principles of Econometrics. New York: Wiley.

Thompson, Warren S. 1930. Danger Spots in World Population. New York: Alfred. A. Knopf.

Thuenen, Johann Heinrich von. 1930. Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie. 3. Jena, Germany: Fischer.

Tournemaine, Frederic, and Pongsak Luangaram. 2012. "R&D, Human Capital, Fertility, and Growth." Journal of Population Economics, 923–953.

Turgot, Anne Robert Jacques. 1767. Réflexions sur la formation et la distribution des richesses.

U.S. Census Bureau. 2011. Statistical Abstract of the United States. Edited by US Department of Commerce, Economics and Statistics Administration. U.S. Census Bureau.

Wallace, Alfred Russel. 1890. Human Selection. Alfred Russel Wallace Classic Writings. 2010.

Ward, Eric J., Kim Parsons, Elizabeth E. Holmes, Ken C. Balcomb, and John Kb Ford. 2009. "The Role of Menopause and Reproductive Senescence in a Long-Lived Social Mammal." Frontiers in Zoology, 4.

Weiner, Jonathan. 1995. The Beak of the Finch. New York.

Wicksteed, P. H. 1914. "The Scope and Method of Political Economy in the Light of the "Marginal" Theory of Value and of Distribution." The Economic Journal, 1–23.

Wicksteed, Philip Henry. 1932. An Essay on the Co-ordination of the Laws of Distribution. 12. London: London School of Economics.

World Bank. 2018. World development indicators. https://datacatalog.worldbank.org/dataset/world-development-indicators.

Wrigley, Edward Anthony, and Roger S. Schofield. 1981. The Population History of England 1541-1871. Cambridge University Press.

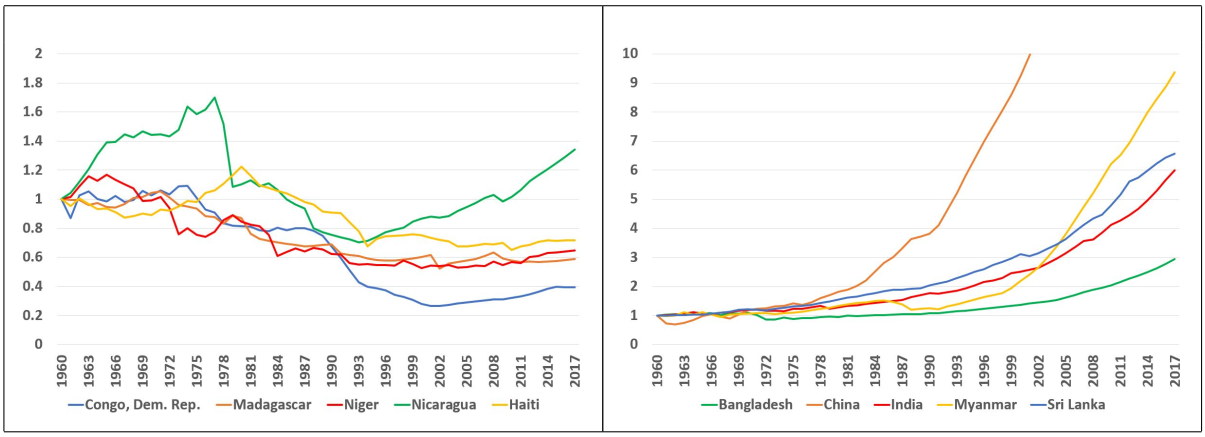
Young, Allyn A. 1928. "Increasing Returns and Economic Progress." The Economic Journal, 527–542.

Chapter 11

Appendix

11.1 Appendix for chapter 2

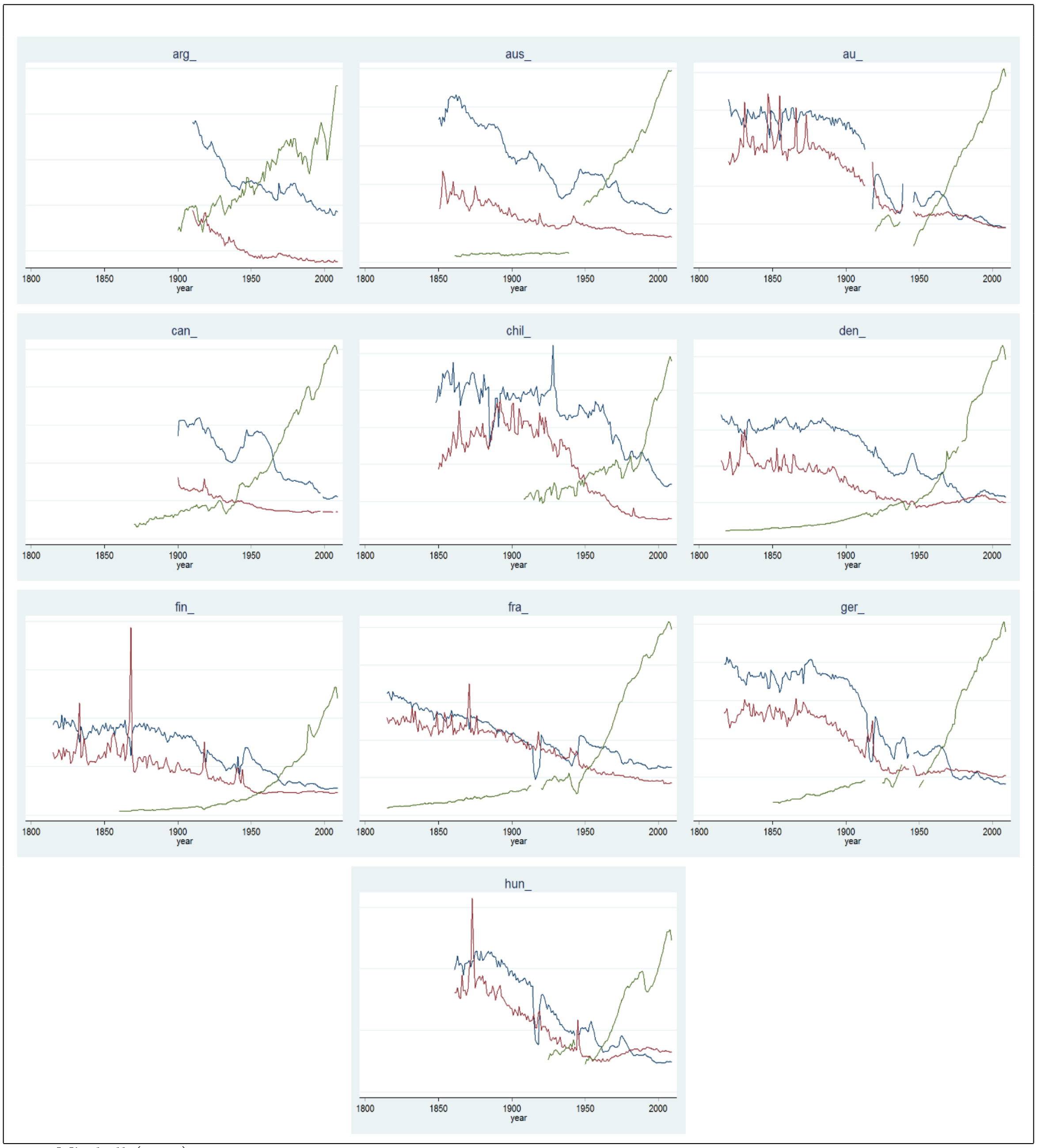
Figure A1: Indexed GDP per capita in selected countries 1960–2017. Left graph: Countries caught in stagnation. Right graph: Countries that have escaped from stagnation.



Source: Worldbank (2018).

11.2 Appendix for chapters 3, 4, 5, 6

Figure B1: 20 selected countries: Birth rates (blue), death rates (red) and GDP per capita (green); GDP per capita is indexed to the year 2010 = 0.05, x- and y-axis intersect at value zero; arg–hun.



Source: Mitchell (2013).

Figure B2: 20 selected countries: Birth rates (blue), death rates (red) and GDP per capita (green); GDP per capita is indexed to the year 2010 = 0.05, x- and y-axis intersect at value zero; ita–ukc.

Source: Mitchell (2013).

11.3 Appendix for chapter 7

11.3.1 Selection criteria

Table C1: British data: Lag selection criteria.

Selection-order criteria						Number of obs = 459 Sample: 1552-2010		
lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-4664.97				136775	20.34	20,350	20,367
1	-3622.06	2085.8	9	0.000	1511.8	15,835	15,877	15,943
2	-3564.38	115.36	9	0.000	1222.85	15,623	15,697	15,812
3	-3533.78	61,203	9	0.000	1113.02	15,528	15,635	15.798*
4	-3506.32	54,917	9	0.000	1027.03	15,448	15.586*	15,799
5	-3496.7	19,242	9	0.023	1024.3	15,445	15,615	15,877
6	-3486.65	20,103	9	0.017	1019.67	15,441	15,643	15,954
7	-3474.39	24,521	9	0.004	1005.37	15,427	15,660	16,020
8	-3460.36	28,052	9	0.001	983,691	15,405	15,670	16,079
9	-3442.45	35,812	9	0.000	946,374	15,366	15,663	16,122
10	-3429.8	25.315*	9	0.003	931.563*	15.345*	15,679	16,187

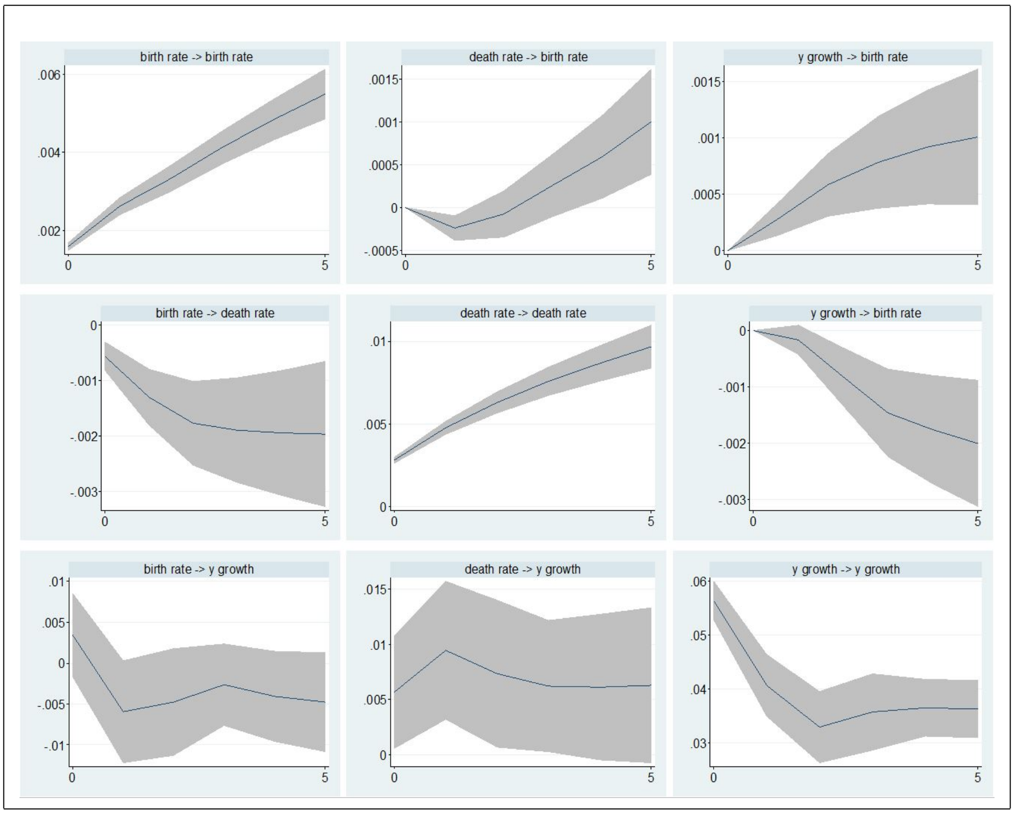
11.3.2 List of countries studied

Table C2: List of twenty countries studied.

CC	Country		CC	Country		CC	Country
arg	Argentina		fra	France		nz	New Zealand
au	Austria		ger	Germany		rom	Romania
aus	Australia		hun	Hungary		spa	Spain
can	Canada		ita	Italy		swe	Sweden
chil	Chile		jap	Japan		swi	Switzerland
den	Denmark		net	Netherlands		ukc	Britain (Clark)
fin	Finland		nor	Norway		ukb	Britain (Broadb.)

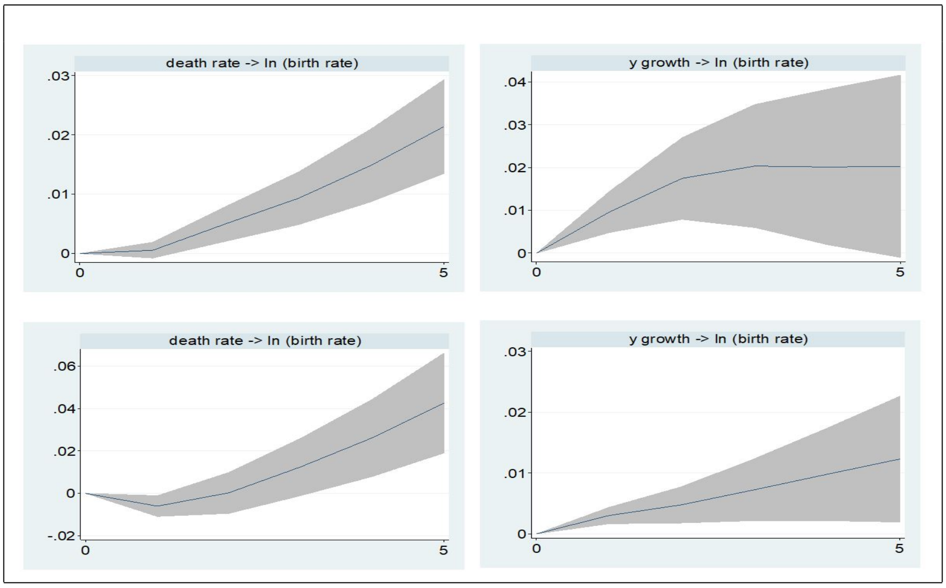
11.3.3 Robustness I: Coirfs using Broadberry et al. data

Figure C1: Britain 1541–2010: Coirf matrix based on a VAR(3) model.



11.3.4 Robustness II: Coirfs of logarithmized birth rates

Figure C2: Coirf of the effects of the PoG (left) and the PoP (right) based on a VAR(3) model using logarithmized birth rates. First line stacked data of 55 economies, second line British historical data.



11.3.5 Robustness III: Coirfs using different variable orderings

Figure C3: Coirf of the effects of the PoG (left) and the PoP (right) based on a VAR(3) model using alternative variable orderings. First line stacked data of 55 economies, second line British historical data. Ordering $(\ b \quad g_y \quad d\ )'$.

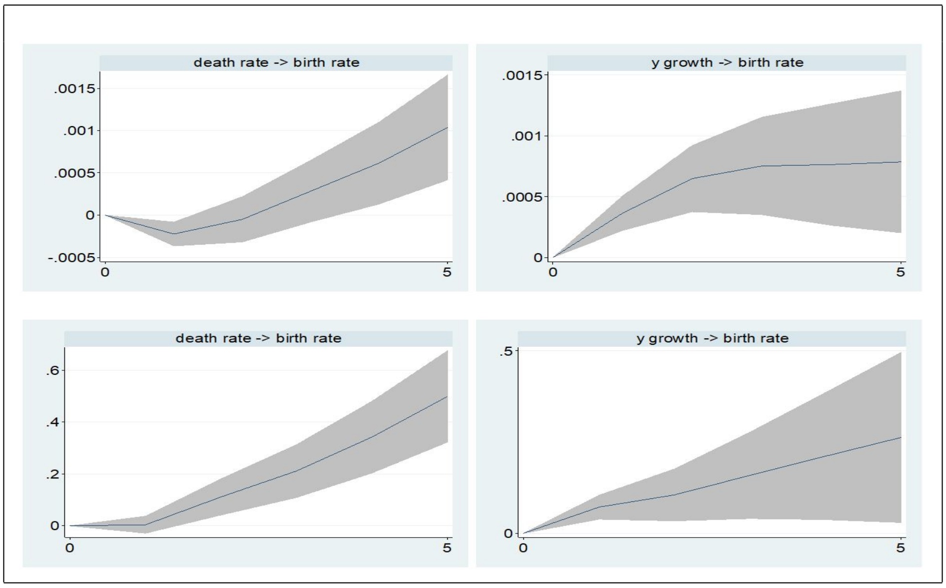


Figure C4: Ordering $(\ g_y \quad b \quad d\ )'$.

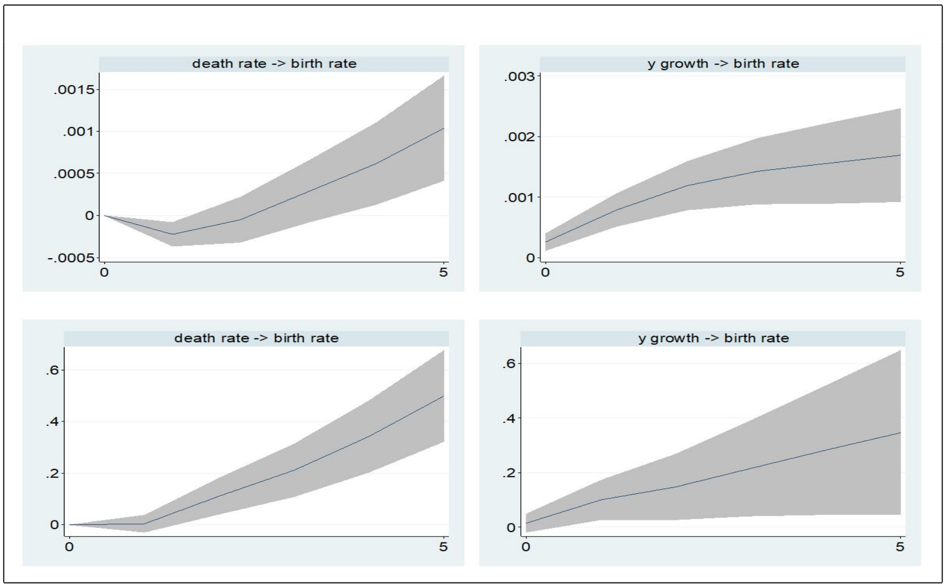


Figure C5: Ordering (g_y d b)$'$.

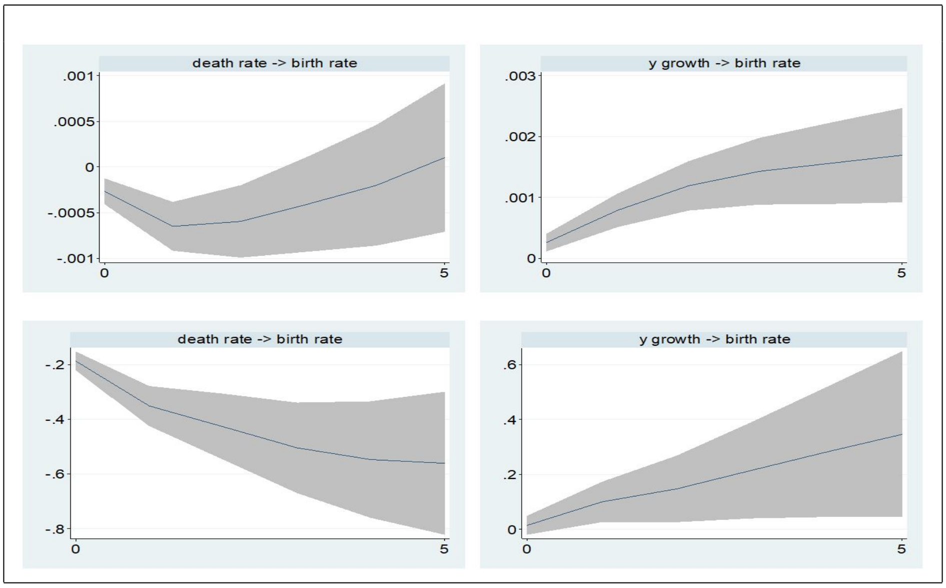


Figure C6: Ordering (d b g_y)$'$.

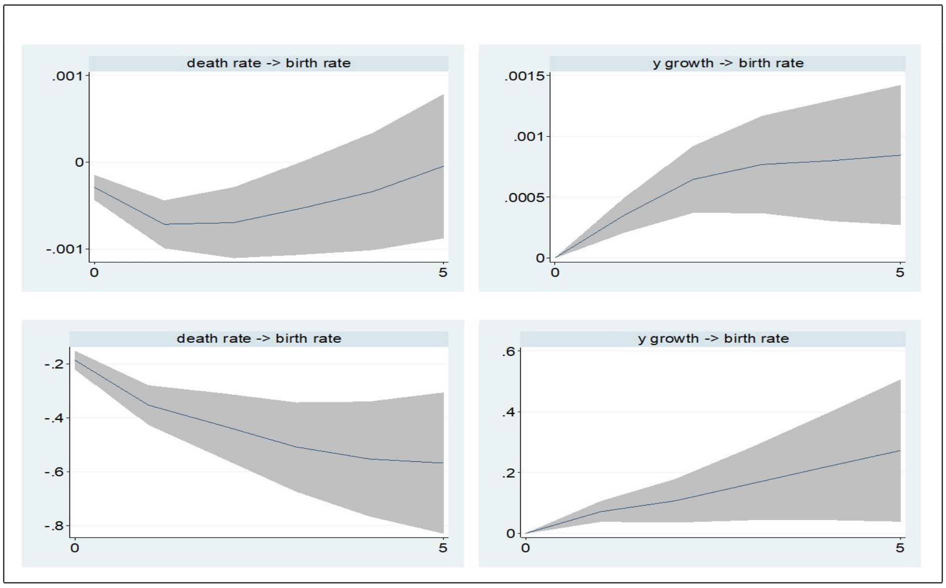
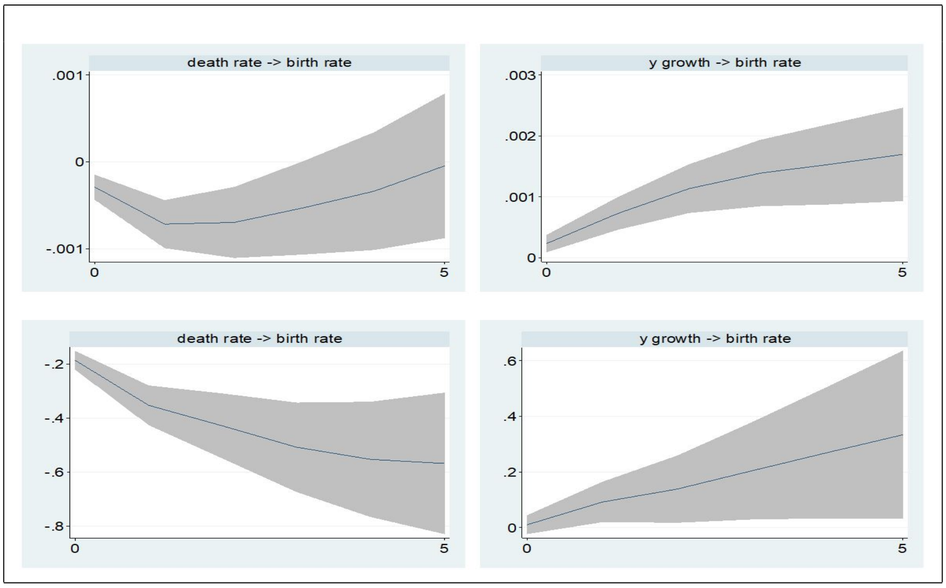
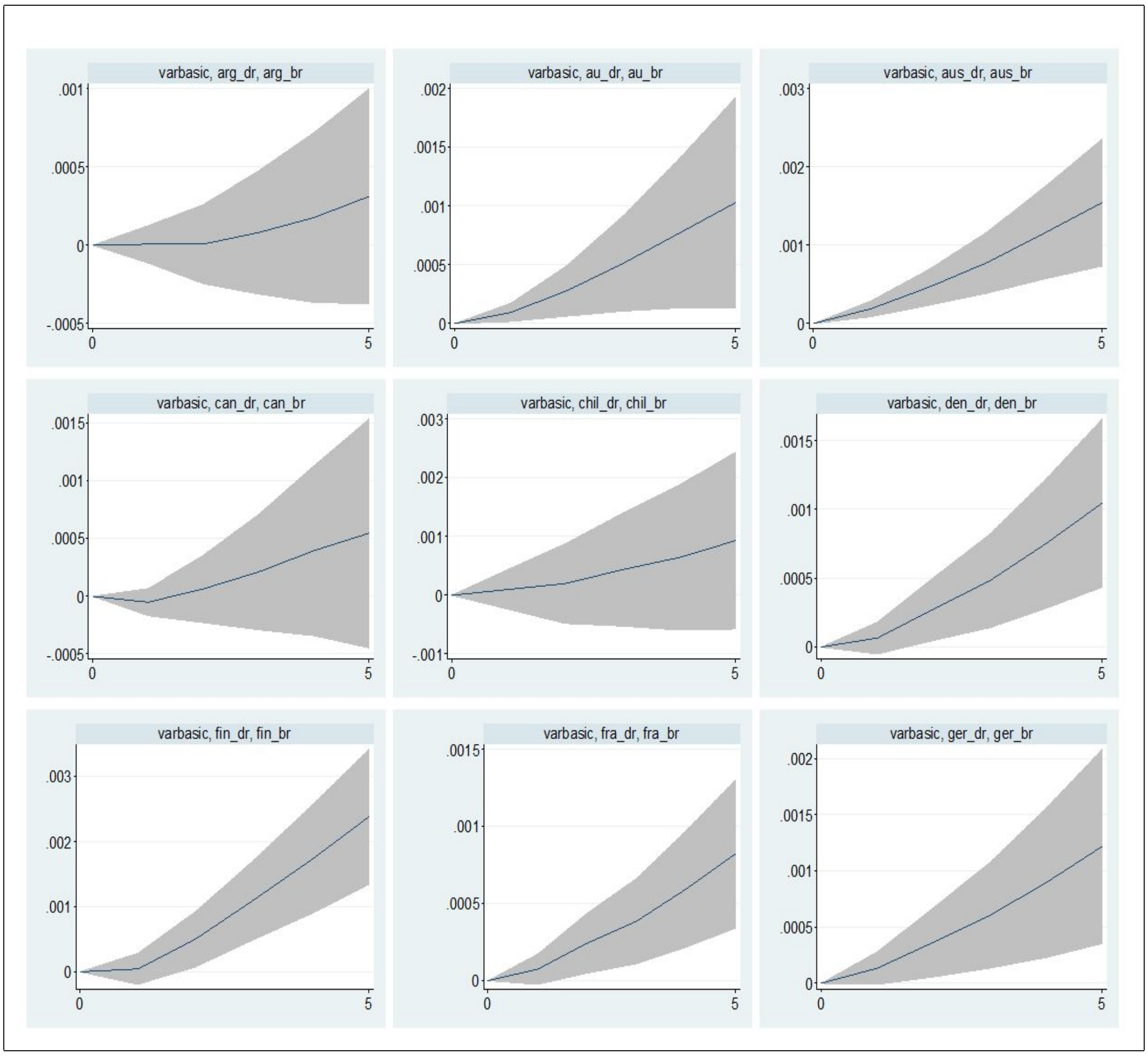


Figure C7: Ordering (d g_y b)$'$.



11.3.6 Robustness IV: Country–specific coirfs

Figure C8: 21 samples: Coirfs of birth rate on a one standard deviation shock in death rate, measuring the PoG.



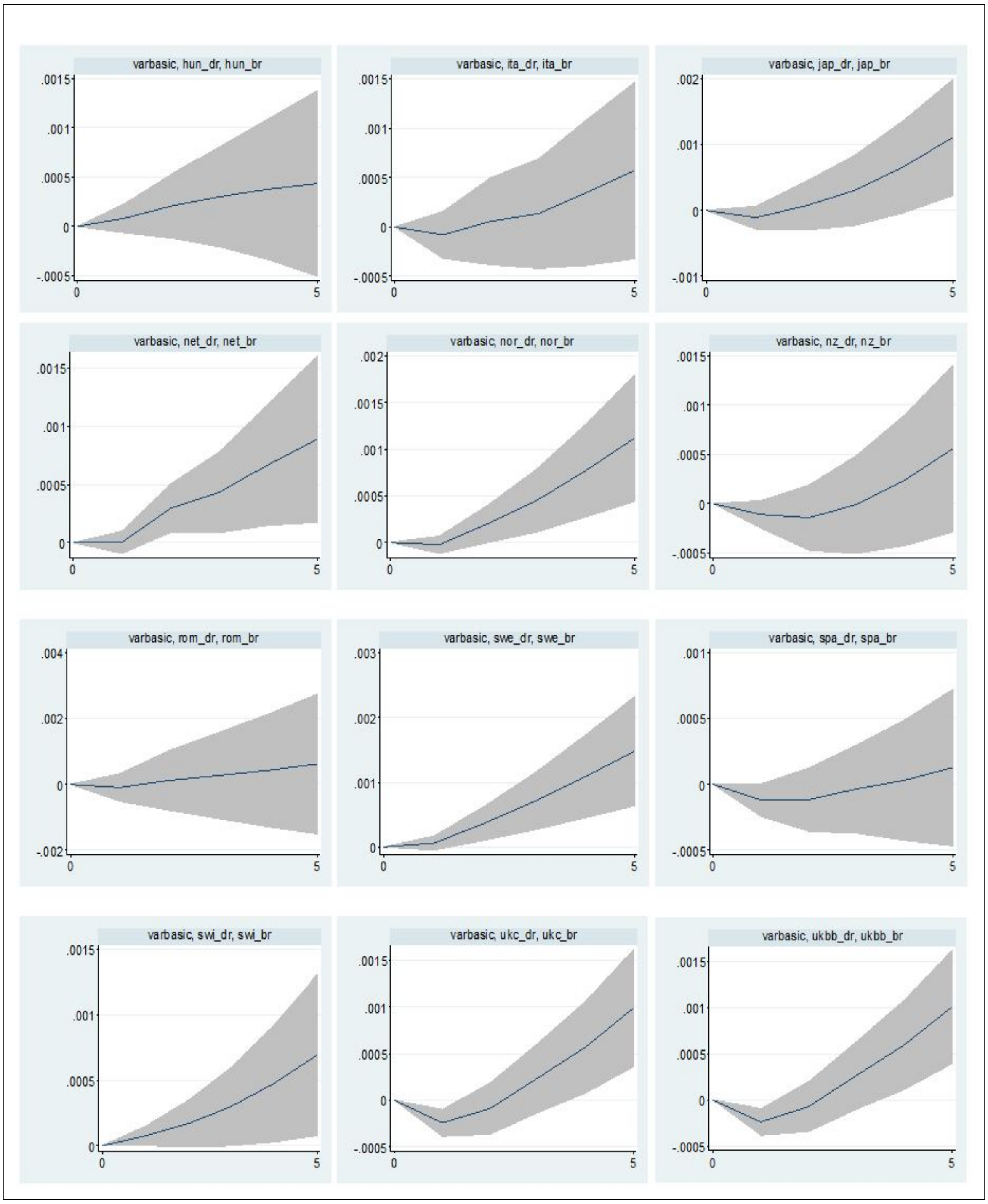
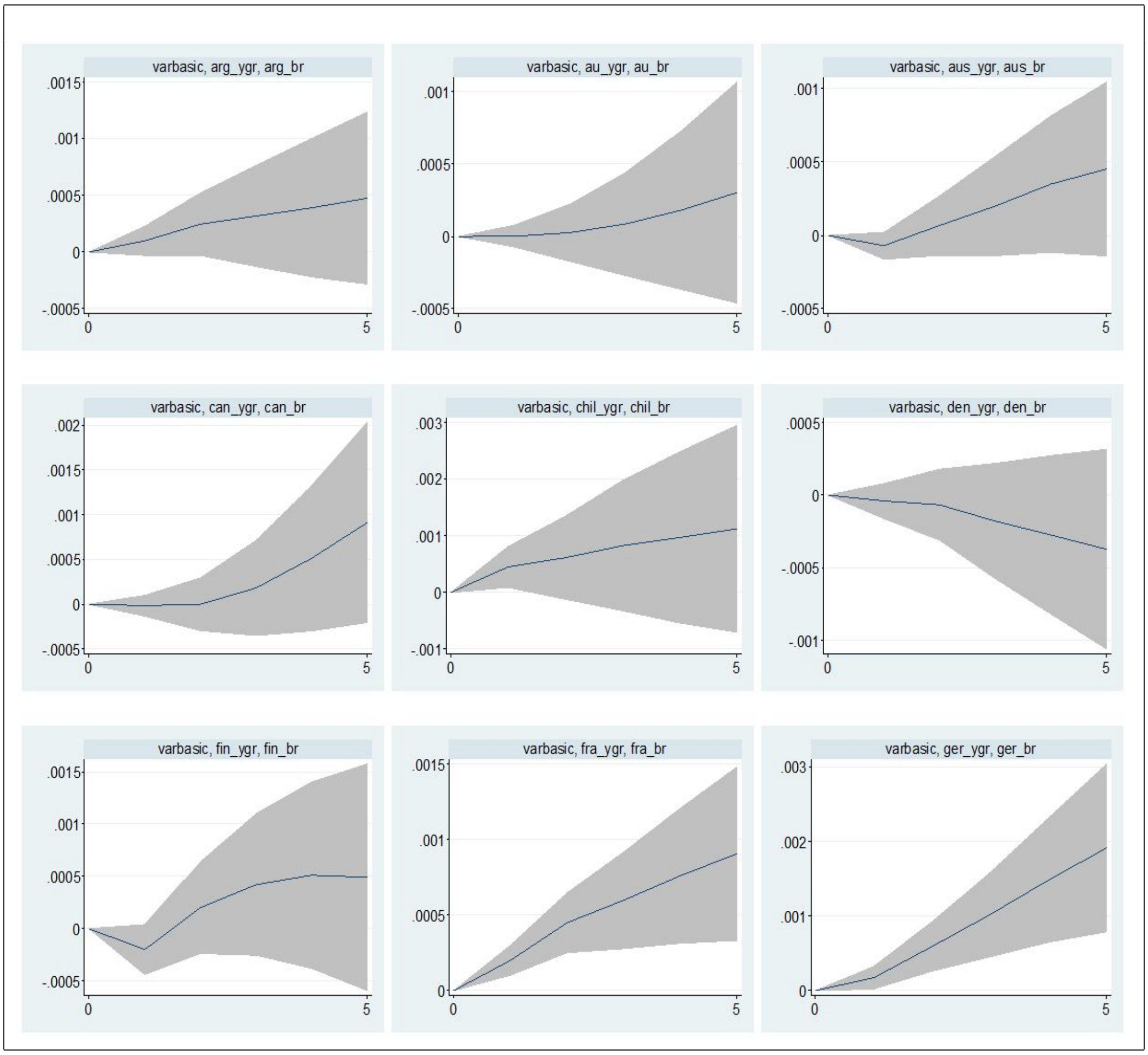
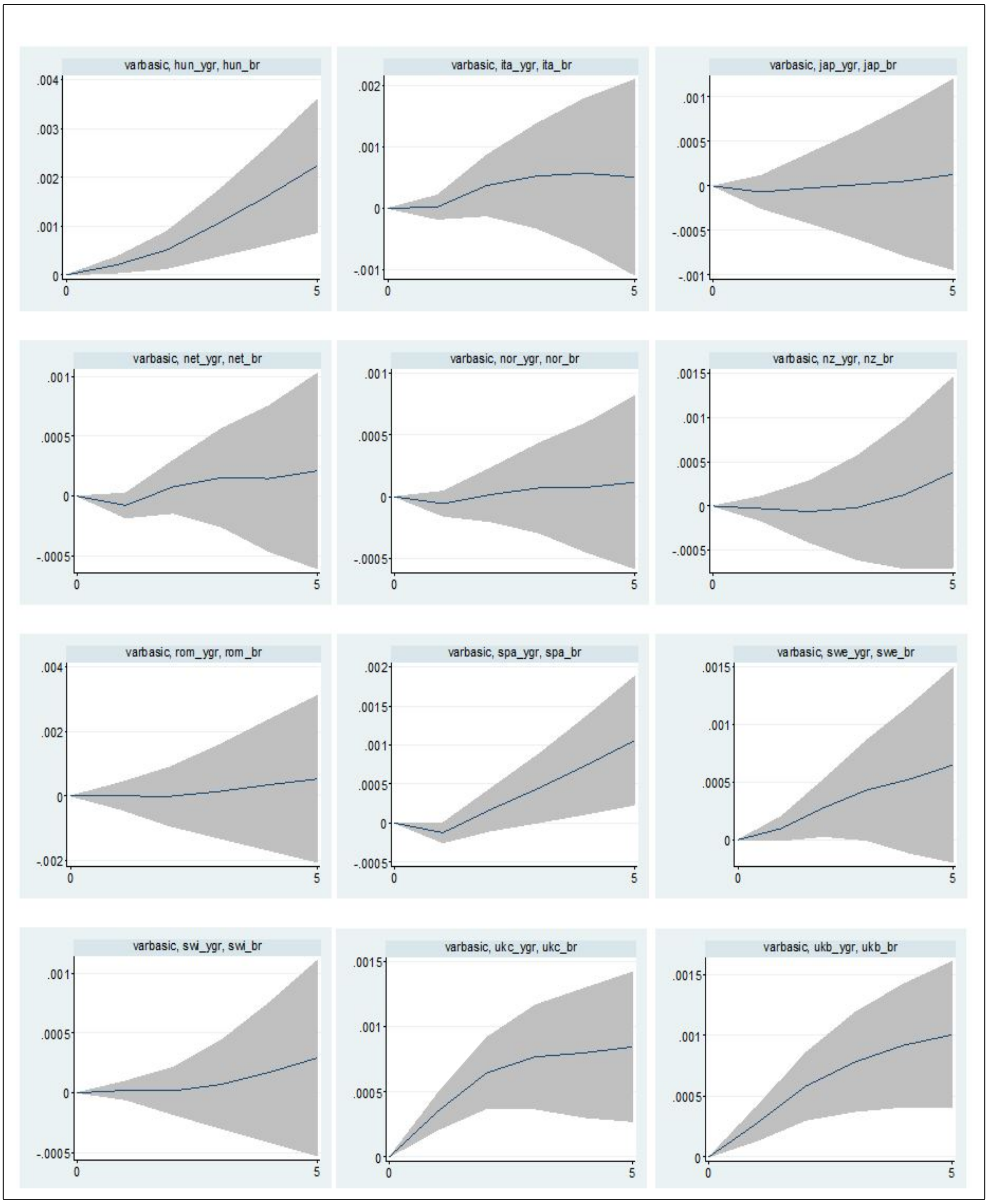

varbasic, hun_dr, hun_br
varbasic, ita_dr, ita_br
varbasic, jap_dr, jap_br
varbasic, net_dr, net_br
varbasic, nor_dr, nor_br
varbasic, nz_dr, nz_br
varbasic, rom_dr, rom_br
varbasic, swe_dr, swe_br
varbasic, spa_dr, spa_br
varbasic, swi_dr, swi_br
varbasic, ukc_dr, ukc_br
varbasic, ukbb_dr, ukbb_br
.0015
.001
.0005
0
-.0005
.002
-.001
.004
-.002
.003
5

Figure C9: 21 samples: Coirfs of birth rate on a one standard deviation shock in GDP per capita growth, measuring the PoP.

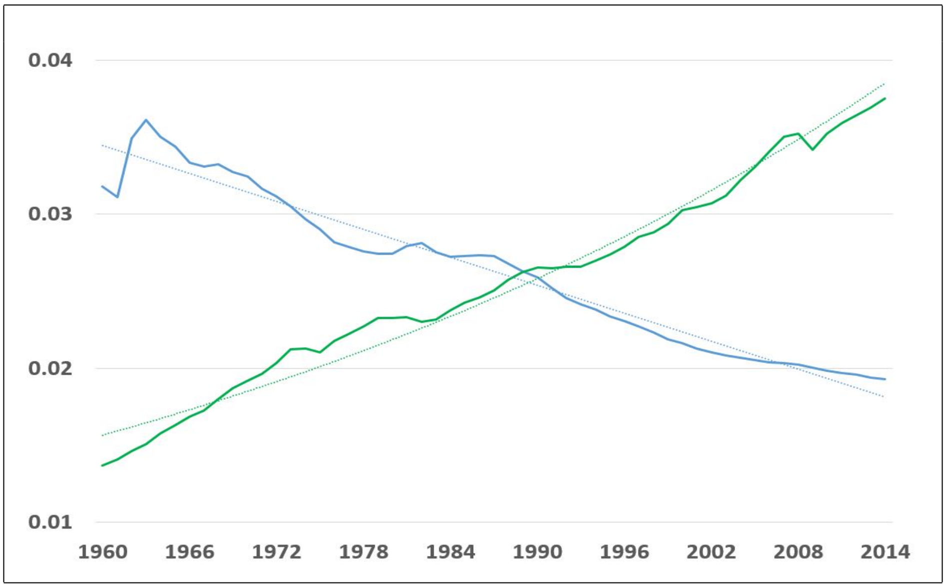



varbasic, hun_ygr, hun_br
varbasic, ita_ygr, ita_br
varbasic, jap_ygr, jap_br
varbasic, net_ygr, net_br
varbasic, nor_ygr, nor_br
varbasic, nz_ygr, nz_br
varbasic, rom_ygr, rom_br
varbasic, spa_ygr, spa_br
varbasic, swe_ygr, swe_br
varbasic, swi_ygr, swi_br
varbasic, ukc_ygr, ukc_br
varbasic, ukb_ygr, ukb_br

11.4 Appendix for chapter 8

11.4.1 Global "cross of wealth"

Figure D1: Birth rate (blue) and indexed GDP per capita (green), aggregate global data, 1960–2014.



Source: World Bank (2018).

11.4.2 List of countries studied

Table D1: List of 104 countries studied. 34 countries with long-run data are starred.

Algeria	Germany	Nicaragua
Argentina*	Ghana	Niger
Australia*	Greece*	Nigeria
Austria*	Guatemala	Norway*
Bangladesh	Guinea	Pakistan
Belgium*	Guinea–Bissau	Panama*
Benin	Haiti	Paraguay
Bolivia	Honduras*	Peru
Botswana	Hong Kong	Philippines*
Brazil	India*	Portugal*
Burkina Faso	Indonesia	Romania
Burundi	Iran, Islamic Rep.	Rwanda
Cameroon	Ireland*	Senegal
Canada*	Israel*	Sierra Leone
Central African Republic	Italy*	Singapore
Chad	Jamaica*	South Africa
Chile*	Japan*	Spain*
China	Jordan	Sri Lanka*
Colombia*	Kenya	Sweden*
Congo, Dem. Rep.	Korea, Rep.	Switzerland*
Congo, Rep.	Lesotho	Syrian Arab Republic
Costa Rica	Madagascar	Tanzania
Cote d'Ivoire	Malawi	Thailand
Cyprus*	Malaysia	Togo
Denmark*	Mali	Trinidad and Tobago
Dominican Republic	Mauritania	Tunisia
Ecuador	Mauritius	Turkey
Egypt, Arab Rep.*	Mexico*	Uganda
El Salvador*	Morocco	United Kingdom*
Equatorial Guinea	Mozambique	United States
Ethiopia	Myanmar	Uruguay*
Finland*	Namibia	Venezuela, RB*
France	Nepal	Zambia
Gabon	Netherlands*	Zimbabwe
Gambia, The	New Zealand*	